Mediating in The Era of Polarization

AGENDA

MEDIATING IN THE ERA OF POLARIZATION

FRIDAY, APRIL 24, 2020

12:00PM	WELCOME AND INTRODUCTION PROFESSOR ALEXANDRA CARTER, COLUMBIA LAW SCHOOL
12:05PM	INTRODUCTION TO POLARIZATION
12:10PM	ETHICAL STANDARDS OVERVIEW
12:15PM	TRANSFERENCE
12:20PM	DECISION POINTS
12:45PM	DISCUSSION
12:55PM	THANK YOU AND FEEDBACK

PROFESSOR ALEXANDRA CARTER

Alexandra Carter is a Clinical Professor of Law and the Director of the Mediation Clinic at Columbia Law School. From 2012-2016, she also served as the Law School's Director of Clinical Programs, providing leadership in the area of experiential teaching and learning.

Professor Carter's teaching, research and publications lie in the field of alternative dispute resolution, primarily in mediation and negotiation. She has been sought as a trainer on mediation, civil procedure, negotiation and dispute systems design for many different groups from private and public sectors, including the United Nations, U.S. courts and federal agencies, private corporations and law firms. She has addressed the Chinese Academy of Social Sciences Rule of Law Conference, the 5th World Peace Conference in Jakarta, the ICU-CLS Peace Summit in Tokyo, and the Ceará Supreme Court Conference on Mediation; and has contributed as a faculty speaker at universities in South America, Asia and Europe. She has served on the Alternative Dispute Resolution Committee for the New York City Bar Association, as well as the Mediator Ethics Advisory Committee for the New York State Unified Court System.

Prior to joining the Columbia faculty, Professor Carter was associated with Cravath, Swaine & Moore LLP, where she worked as part of a team defending against a multibillion dollar securities class action lawsuit related to the Enron collapse, served as the senior antitrust associate on several multibillion dollar mergers, and handled cases involving copyright law.

Professor Carter received her Juris Doctor degree in 2003 from Columbia Law School, where she earned James Kent and Harlan Fiske Stone academic honors. She also won the Jane Marks Murphy Prize for clinical advocacy and the Lawrence S. Greenbaum Prize for the best oral argument in the 2002 Harlan Fiske Stone Moot Court Competition. After earning her degree, Professor Carter clerked for the Hon. Mark L. Wolf, U.S. District Court for the District of Massachusetts in Boston.

PROFESSOR CARTER'S RESEARCH TEAM

JENNIFER ANGE
Jennifer Ange is a third year student at Columbia Law School. She is a member of the Mediation Clinic, the Columbia Law Women's Association, and the Asian Pacific American Law Student Association. She is also a Submission Editor for the Columbia Science and Technology Law Review.

JAMIE CHANG
Jamie Chang is a third year student at Columbia Law School. She is a member of the Mediation Clinic, the Business and Law Association, and the Asian Pacific American Law Student Association.

AUDREY CHAO
Audrey Chao is a third year student at Columbia Law School and second year student at Columbia School of Social Work. She is a member of the Mediation Clinic and a Managing Editor of Columbia's Human Rights Law Review.

HENRY GUERRA
Henry Guerra is a second year student at Columbia Law School. He is a member of the Latinx Law Student Association and First-Generation Professionals, participates in the Mediation Clinic, and served as a staffer for the Columbia Human Rights Law Review and Jailhouse Lawyer's Manual.

OLIVER KWON
Oliver Kwon is a second year student at Columbia Law School. He is a member of the Mediation Clinic, the Real Estate Law Society, and the Asian Pacific American Law Student Association.

HALEY LING
Haley Ling is a third year student at Columbia Law School. She is a member of the Mediation Clinic, the Columbia Health Law Association, and the Asian Pacific American Law Student Association. She is also Development Editor of the Columbia Law Review.

MELISSA MEZA
Melissa Meza is a third year student at Columbia Law School. She is a member of the Mediation Clinic, the Latinx Law Student Association, Empowering Women of Color, and the Society for Immigrant and Refugee Rights.

ESTHER PORTYANSKY
Esther Portyansky is a third year student at Columbia Law School. She is a member of the Mediation Clinic, the Columbia Law Women's Association, the Jewish Law Students' Association, and the Columbia Law Revue.

About the New York State Dispute Resolution Association (NYSDRA)

NYSDRA is a nationally-recognized association of members including the Community Dispute Resolution Centers (CDRC) network, individuals, students and organizations. We provide dispute resolution services and training in communities throughout New York State.

The New York State Dispute Resolution Association is a statewide resource and advocacy organization, providing conflict resolution program and contract management, technical assistance, training, advocacy, outreach, education, and professional networking opportunities to ensure quality conflict management and peaceful dispute resolution.

NYSDRA Membership & Benefits

NYSDRA membership is essential for both experienced and aspiring ADR practitioners and providers in New York. As part of your membership, you will receive access to key resources, advocacy and promotion, and practical support. Join this dynamic community and connect with the network today. Member benefits include:

Access & Resources
- Live Members only webinars, CLE programs, & online webinar archive
- *Conflict Resolution Quarterly*

Advocacy & Promotion
- Marketing support & promotion for you
- NYSDRA lobbies NYS government officials to advocate for and advance the field of dispute resolution.
- Dedicated NYSDRA website profile & dashboard – share your ADR profile and network with the NYSDRA community

Practice & Technical Support and Networking Opportunities
- Access to professional liability insurance
- Exclusive trainings
- Discounted conference registration & free regional networking events
- Consulting with professional development network peers
- Guidance navigating NYS ADR landscape / practice development

For more information or to become a member today

visit **NYSDRA.org** or connect with us at **info@nysdra.org**

New York State Dispute Resolution Association, Inc.
4 Pine West Plaza, Suite 411 – Albany, New York 12205

CITATIONS

I. TRANSFERENCE

1. James Fauth & Jeffrey A. Hayes, *Counselors' Stress Appraisals as Predictors of Countertransference Behavior With Male Clients*, 64 J. Counseling & Dev. 430 (2006).

2. Prasko et al., *Transference and Countertransference in Cognitive Behavioral Therapy*, 154 Biomedical Papers Med. Fac. U. Palacky, Olumouc, Czech 189 (2010).

3. Charles Gelso & Jonathan Mohr, *The Working Alliance and the Transference/Countertransference Relationship: Their Manifestation with Racial/Ethnic and Sexual Orientation Minority Clients and Therapists*, 10 Applied & Preventive Psychol. 51 (2001).

4. Steven Tublin, *Political Identity and Countertransference*, *in* Unknowable, Unspeakable, and Unsprung: Psychoanalytic Perspectives on Truth, Scandal, Secrets, and Lies 71 (Jean Petrucelli & Sarah Schoen eds., 2016)

II. CASE DEVELOPMENT

1. Mark D. Alicke, Frank M. Loschiavo & Justin T. Buckingham, *Attitude Transference*, 18 Soc. Cognition 1 (2000).

2. Matthew P. Deegan et al, *Positive Expectations Encourage Generalization from a Positive Intergroup Interaction to Outgroup Attitudes*, 4 Personality & Soc. Psychol. Bull. 52 (2014).

3. Susan T. Fiske, Amy J.C. Cuddy & Peter Glick, *Universal Dimensions of Social Cognition: Warmth & Competence*, 11 Trends Cognitive Sci. 77 (2007).

4. Rajiv Jhangiani & Hammond Tarry, *Initial Impression Formation*, Principles of Social Psychology (1st int'l ed. 2014), https://opentextbc.ca/socialpsychology/chapter/initial-impression-formation/.

5. Matthew Rabin & Joel L. Schrag, *First Impressions Matter: A Model of Confirmatory Bias*, 114 Q. J. Econ. 37 (1999).

6. John Paul Wilson & Nicholas O. Rule, *Perceptions of Others' Political Affiliation are Moderated by Individual Perceivers' Own Political Attitudes*, 9 Pub. Libr. Sci. ONE 1 (2014).

III. SUMMARIZING

1. Erik Cleven et al., *Living with No: Political Polarization and Transformative Dialogue*, 2018 J. Disp. Resol. 53 (2018).

2. Lindsey Phipps, *A Divided Nation: Political Polarization and Dispute Resolution*, 17 Pepp. Disp. Resol. L.J. 111 (2017).

3.   Carol Pauli, *Transforming News: How Mediation Principles Can Depolarize Public Talk*, 15 Pepp. Disp. Resol. L.J. 85 (2015).

4.   Lee De-Wit et al., *What Are the Solutions to Political Polarization?*, Greater Good (July 2, 2019).

IV.   REFRAMING

1.   John M. Livingood, *Reframing and Its Uses*, 57 J. Dispute Resolution 44 (2002).

2.   Deborah Shmueli, Michael Elliott & Sanda Kaufman, *Frame Changes and the Management of Intractable Conflicts*, 24 Conflict Resolution Q. 207 (2006).

3.   Jan G. Voelkel & Matthew Feinberg, *Morally Reframed Arguments Can Affect Support for Political Candidates*, 9 Soc. Psych & Personality Sci. 917 (2018)

4.   Matthew Feinberg & Robb Willer, *Moral Reframing: A Technique for Effective and Persuasive Communication Across Political Divides*, Soc. & Personality Psychol. Compass (2019).

V.   ACKNOWLEDGEMENT AND STROKING

1.   Carol Izumi, *Implicit Bias and the Illusion of Mediator Neutrality*, 34 Wash. U. J. L. & Pol'y 71 (2010).

2.   Ginsey Varghese Kramarczyk, *The Media, a Polarized America & ADR Tools to Enhance Understanding of Perspectives*, 19 Pepp. Disp. Resol. L.J. 127 (2019).

3.   Carrie Menkel-Meadow, *Why We Can't Just All Get Along: Dysfunction in the Polity and Conflict Resolution and What We Might Do about It*, 2018 J. Disp. Resol. 5 (2018).

4.   Robert C. Bordone, *Building Conflict Resilience: It's Not Just about Problem-Solving*, 2018 J. Disp. Resol. 65 (2018).

VI.   AGENDA SETTING

1.   Erick Cleven et al., *Living With No: Political Polarization and Transformative Dialogue*, Hofstra J. Disp. Resol. 53, 60-63 (2018).

2.   Howard S. Bellman & Susan L. Podziba, *Public Policy Mediation*, 18 Disp. Resol. Mag. 14 (2012).

3.   Carol Izumi, *Implicit Bias and the Illusion of Mediator Neutrality*, 34 Wash. U. J.L. & Pol'y 71 (2010).

4.   Lindsey Phipps, *A Divided Nation: Political Polarization and Dispute Resolution*, 17 Pepp. Disp. Resol. L.J. 111 (2017).

VII. CAUCUSING

1. Gregorio Billikopf-Encina, *Contributions of Caucusing and Pre-Caucusing to Mediation*, 4 Group Facilitation 3 (2002)

2. Alan Gross, *Raising Awareness of Potential Biases and Microaggressions: Applications to Mediation*, 3 J. Mediation and Applied Conflict Analysis 407 (2016).

3. Gary Welton, Dean Pruitt & Neil McGillicuddy, *The Role of Caucusing in Community Mediation*, 32 J. Conflict Resolution 181 (1988).

4. Jean Poitras, *The Strategic Use of Caucus to Facilitate Parties' Trust in Mediators*. 24 Int'l J. Conflict Mgmt. No. 1 (2013).

VIII. POLARIZATION

1. Cass R. Sunstein, *The Law of Group Polarization*, 10 J. Pol. Phil. 175 (2002).

2. Cass R. Sunstein, *Group Polarization and* 12 Angry Men, 23 Negot. J. 443 (2007).

3. Robert B. Talisse, *Political Polarization Is About Feelings, Not Facts*, The Conversation (July 31, 2019), http://theconversation.com/political-polarization-is-about-feelings-not-facts-120397.

Counselors' Stress Appraisals as Predictors of Countertransference Behavior With Male Clients

James Fauth and Jeffrey A. Hayes

The authors investigated the applicability of a transactional theory of stress to the understanding of countertransference with male clients. Counselors responded to either a traditional or nontraditional videotaped male client vignette. As expected, counselors' stress appraisals predicted their countertransference behavior. Specifically, therapists' positive appraisals were linked with more positive diagnostic evaluations of the client, whereas negative appraisals were linked with increased distance from and hesitance with the client. Implications of these results are discussed.

Countertransference is a double-edged sword for counselors and researchers alike. In counseling, when countertransference is unchecked, it can lead to countertherapeutic behavior (Gelso & Hayes, 1998; Strupp, 1980); conversely, if counselors are aware of countertransference, it can provide crucial insight into the nature of the counseling relationship (Singer & Luborsky, 1977). For researchers, conducting countertransference research is difficult due to definitional and methodological difficulties (Gelso & Hayes, 1998). Nevertheless, because the counselor's and the client's personal reactions and the manner in which they deal with these reactions are central to counseling effectiveness (Strupp, 1980), countertransference research is of critical importance.

Freud introduced the term *countertransference* (Freud, 1910/1959) to refer to the analyst's unconscious and neurotic reactions to the patient's transference. Subsequent writers broadened Freud's definition of countertransference to include all of a counselor's reactions to a client (Fromm-Reichmann, 1950; Kiesler, 1982; Levenson, 1995). Currently, many scholars distinguish between therapists' "real" and "unreal" reactions, defining countertransference as irrational reactions emanating from counselors' unresolved internal issues (Gelso & Carter, 1985, 1994; Gelso & Hayes, 1998). In the present study, we used this current view of countertransference, which retains the neurotic element of Freud's definition without limiting countertransference strictly to unconscious reactions in response to transference.

The Current Countertransference Research Paradigm

In this section, we briefly describe the current countertransference research paradigm, focusing on three countertransfer-

ence components (i.e., countertransference origins, triggers, and manifestations) identified by Hayes (1995) that are applicable to the current investigation. We have included this review for two primary reasons: (a) to highlight the empirical basis for the current investigation and (b) to serve as the historical context from which our reconceptualization of countertransference has emerged.

Within the current countertransference paradigm, counselors' unresolved issues (i.e., countertransference origins) are thought to interact with client characteristics (i.e., countertransference triggers) to elicit counselors' cognitive, affective, and/or behavioral reactions (i.e., countertransference manifestations; Hayes, 1995). To date, research has focused on counselors' intrapersonal (Hayes & Gelso, 1991; Yulis & Kiesler, 1968) and interpersonal (Cutler, 1958; Fauth, Hayes, Park, & Friedman, 1999; Rosenberger & Hayes, 2002) countertransference origins, with increasing attention being paid to countertransference origins of a cultural nature, such as homophobia (Gelso, Fassinger, Gomez, & Latts, 1995; Hayes et al., 1998; Hayes & Gelso, 1993; Latts & Gelso, 1995). These studies typically hypothesized that therapists' unresolved issues (i.e., countertransference origins) and clients' characteristics (i.e., countertransference triggers) would interact to elicit therapists' countertransference manifestations, such as withdrawal from (e.g., avoidance behavior) or overinvolvement with (e.g., enmeshed behavior) clients (Hayes et al., 1998). This notion that an interaction between counselors' and clients' unresolved issues triggers counselors' countertransference was partially supported in four of six studies in which it was tested (Cutler, 1958; Fauth et al., 1999; Gelso et al., 1995, Hayes & Gelso, 1991,

1993; Rosenberger & Hayes, 2002). For instance, Gelso et al. (1995) found that counselors' homophobia levels were directly related to increased avoidance behavior with lesbian clients but not with heterosexual clients.

Countertransference as a Stress and Coping Process

The prevailing countertransference research paradigm has been successful in promoting increased empirical research. Clearly, when counselors' countertransference origins are triggered in some way by clients' characteristics, it can produce countertransference manifestations such as avoidance behavior. It is also clear, however, that such a match between counselors' and clients' unresolved issues sometimes does not elicit counselors' countertransference manifestations.

We believe that the current paradigm is lacking in that it places the therapist in an artificially passive and reactive role within a unidirectional (i.e., from client to therapist) context, thereby artificially obscuring counselors' active role in perceiving, shaping, and coping with the countertransference process. We have addressed this problem by reformulating countertransference in accord with Lazarus and colleague's (Lazarus, 1993; Lazarus & Folkman, 1984) transactional theory of stress. Conceptualizing countertransference in this manner lends itself to viewing the therapist as an active appraiser of and participant in the therapeutic situation. In transactional theory, stress is defined as a bidirectional person–environment relationship that is appraised as taxing one's resources and endangering one's well-being (Folkman, 1984). The heuristic value of the transactional theory has been widely supported (Folkman & Lazarus, 1985; Stein, Folkman, Trabasso, & Richards, 1997; Tomaka, Blascovich, Kibler, & Ernst, 1997).

Within transactional theory, the personal meaning of an environmental event is based on primary and secondary appraisals. In primary appraisal, the person judges whether the environmental event is stressful. Primary appraisals can be positive or negative in valence. Positive appraisals refer to the opportunity for personal growth and gain and are characterized by positive emotions such as excitement, whereas negative appraisals refer to the potential for or to actual personal loss or damage and are characterized by negative emotions such as anger and sadness (Folkman, 1984; Lazarus, 1993; Lazarus & Folkman, 1984). Primary appraisals have tended to be potent predictors of reactions and adjustment to environmental encounters (Stein et al., 1997; Tomaka et al., 1997).

For stressful events, individuals assess the extent to which their perceived coping abilities, resources, or options are capable of meeting the demands of the situation (Lazarus & Folkman, 1984). Self-efficacy, an individual's judgment of how well she or he can perform the courses of action required to effectively deal with a situation, is a crucial com-

ponent of the secondary appraisal process (Folkman, 1984; Lazarus & Folkman, 1984). Indeed, efficacious judgments decrease emotional arousal and bolster direct coping efforts (Bandura, 1982, 1997).

Thus, we posit that countertransference is a stressful interpersonal event in which the therapist appraises the counseling situation as harmful to, threatening, challenging, and/or taxing of her or his coping resources. On the basis of these appraisals, counselors initiate coping aimed at managing the demands of the situation (thereby potentially creating a reciprocal process). Counselors' subjective appraisals, then, could be the key to understanding the countertransference process, effectively explaining occasions in which therapists' and clients' unresolved issues do not interact to elicit countertransference manifestations and vice versa (occasions in which countertransference manifestations appear in the absence of any interaction between countertransference origins and triggers).

Description of the Study

To provide an initial test of the transactional theory of countertransference, we modeled our study after previous countertransference investigations, with one major exception: We measured counselors' stress appraisals. The inclusion of counselors' stress appraisals allowed us to test two important tenets of transactional theory. The first, and most crucial, tenet was whether counselors' stress appraisals would in fact predict counselors' countertransference manifestations (irrespective of any interaction between countertransference origins and triggers). The second tenet was whether counselors' stress appraisals would mediate any interaction between countertransference origins (i.e., male gender role attitudes) and triggers (i.e., client condition). It is important to note that the transactional theory does not necessarily predict that there should be an interaction effect; it does, however, predict that if such an interaction is found, it should be mediated by (i.e., work through) counselors' stress appraisals in influencing counselors' stress appraisals.

Thus, we examined the relationship between therapists' male gender role attitudes (i.e., countertransference origin); male clients' gender role conformity (i.e., countertransference trigger); therapists' primary and secondary stress appraisals; and therapists' avoidance behaviors, ratings of clients' functioning, and hesitance (i.e., countertransference manifestations). Client verbal content (traditional vs. nontraditional) and nonverbal behavior (traditional vs. nontraditional) were manipulated to create two male client vignettes: (a) a traditional client in terms of both verbal content and nonverbal behavior and (b) a nontraditional client in terms of both verbal content and nonverbal behavior. We manipulated both client nonverbal behavior and verbal content in our videotaped male client vignettes because previous analog research has neglected the importance of clients' nonverbal behavior in triggering countertransference.

We examined counselors' male gender role attitudes (a novel countertransference origin) in conjunction with male clients' gender role conformity (a novel countertransference trigger) because research suggests that counselors sometimes exhibit clinical biases toward male clients based on their own gender role socialization experiences (Fitzgerald & Cherpas, 1985; Robertson & Fitzgerald, 1990; Wisch & Mahalik, 1999). Specifically, Wisch and Mahalik found that counselors experiencing more gender role conflict reacted more negatively toward an angry gay male client vignette than did counselors experiencing less gender role conflict. Furthermore, they found that counselors who were more comfortable expressing affection toward other men reacted more positively toward gay male client vignettes than toward heterosexual male client vignettes.

In measuring counselors' countertransference manifestations, we incorporated two important developments in countertransference research by using multidimensional assessment and a reliable measure of therapist avoidance (Bandura, Lipsher, & Miller, 1960; Gelso et al., 1995; Hayes & Gelso, 1993). We also created the first nonverbal countertransference manifestation measure, namely, counselors' hesitance in responding to clients. Nonverbal measures are crucial because they might capture counselors' unconscious, or peripheral, countertransference behaviors (Gelso & Hayes, 1998; Sherman, 1965).

In sum, our primary purpose in conducting this study was to assess the potential scientific value of the transactional model of countertransference. According to the model, counselors' stress appraisals should be the primary determinants of their countertransference manifestations with clients. Thus, we hypothesized that counselors' stress appraisals should predict their countertransference manifestations. Furthermore, *if* counselors' male gender role attitudes (i.e., countertransference origin) and clients' male gender role conformity (i.e., countertransference trigger) interacted to predict counselors' countertransference manifestations, we hypothesized that the interaction would be *mediated* by counselors' stress appraisals. In other words, according to the transactional model, any such interaction should influence counselors' countertransference manifestations indirectly via its effect on counselors' stress appraisals.

Method

Participants and Selection Procedures

All participants were counselors-in-training or counseling center clinicians from a large eastern university. Potential participants ($N = 154$) were contacted by means of campus mail and asked to volunteer for a counseling process study. Participants were informed that the study included responding to a videotaped client vignette and completing several self-report questionnaires. A raffle for two $100 savings bonds was offered as an incentive to participate.

Sixty-eight counselors agreed to participate, a 44% response rate. Two participants who failed to complete the study were dropped from all analyses. Two thirds of the participants were female ($n = 44$), and one third were male ($n = 22$). Participants' ages ranged from 22 to 53 years ($M = 31.00$, $SD = 8.89$). Most participants were Caucasian ($n = 49$, 74%) or African American ($n = 11$, 17%). A small number of Asian American ($n = 2$), Asian International ($n = 1$), biracial ($n = 1$), Hispanic American ($n = 1$), and Native American ($n = 1$) counselors also participated. Participants' clinical experience varied, with completed practica ranging from 0 to 14 semesters ($M = 3.56$, $SD = 3.28$) and clinical work experience ranging from 0 to 25 years ($M = 4.36$, $SD = 6.55$). A diverse array of theoretical orientations was represented, with humanistic (34%), cognitive-behavioral (20%), psychodynamic (17%), and multicultural (13%) models predominating.

To assess whether counselors-in-training ($n = 38$) and practicing counselors ($n = 28$) differed on any of the variables of interest (i.e., challenge appraisals; negative appraisals; secondary appraisals; avoidance behavior; Global Assessment of Functioning [GAF] Scale, American Psychiatric Association [APA], 2000, ratings; and hesitance scores), a multivariate analysis of variance (MANOVA) was conducted. The MANOVA was not statistically significant, $F(7, 54) = 1.29$, $p > .05$. Thus, the data for counselors-in-training ($n = 38$) and practicing counselors ($n = 28$) were combined for all subsequent analyses. Similarly, we assessed whether male versus female counselors differed on any of the aforementioned measures. Again, the MANOVA was not statistically significant, $F(7, 54) = 1.35$, $p > .05$; therefore, the data for male ($n = 22$) and female ($n = 44$) counselors were combined in all subsequent analyses.

Experimental Conditions and Stimulus Materials

A White male actor in his 30s portrayed a client in videotaped vignettes designed to correspond to the two client conditions. The client, who was shown at close range, faced the camera and spoke as if addressing his therapist. Both vignettes included five segments (each approximately 2.5 minutes in length), with each segment addressing a different aspect of the client's presenting problem. At the end of each segment, a "Please Respond" screen appeared for 20 seconds. This screen cued counselors to pause the tape and make their verbal response to the client.

Male gender role conflict theory guided the construction of the client vignettes. This theory suggests that traditional and nontraditional men differ along four dimensions: restrictive emotionality; restricted affectionate behavior between men; conflict between work and family; and a desire for success, power, and competition (Good et al., 1995; O'Neil, 1980). The traditional male client adhered to these norms: He was unemotional, played sports with "the guys," struggled with work and family values, and strove for success and power. In contrast, the nontraditional male client

departed from these norms: He was openly emotional, enjoyed intimate conversations with male friends, was a stay-at-home dad, and occupied a dependent marital role.

Across the two conditions, the clients discussed parallel problems with marital difficulties and depression. Each of the five segments dealt with different content. First, the traditional male client described his role as the family provider, while the nontraditional client described his role as the primary homemaker. Second, the traditional client was frustrated with his wife's wish that he be more "sensitive"; conversely, the nontraditional client was upset that his wife expected him to be more "macho." Third, the traditional client wanted his wife to be more involved with the children; in contrast, the nontraditional client felt guilty about his own noninvolvement with the children. Fourth, the traditional client was angry over his failed attempts to initiate sex, whereas the nontraditional client was sad about his failed attempts to create more intimacy. Fifth, the traditional client regretted his inability to "draw the line" with his wife, whereas the nontraditional client regretted his dependent behavior with his wife.

The clients' emotional expressiveness was manipulated by varying the nonverbal behaviors of the traditional versus the nontraditional male client vignettes. The traditional client was distant and unemotional. He made infrequent eye contact and leaned slightly away from the therapist. In contrast, the nontraditional client was engaging and emotional. He made frequent eye contact and leaned slightly toward the therapist.

To assess the experimental manipulation as well as the plausibility and believability of the videotaped male client vignettes, we collected data from counselor trainees enrolled in two sections of a counseling theory course ($N = 52$). Students in the first section of the course ($n = 25$) rated the nontraditional male client vignette, while students in the second section of the course ($n = 27$) rated the traditional male client vignette.

To assess the experimental manipulation, participants rated six items on a 7-point Likert-type scale (1 = *strongly disagree,* 7 = *strongly agree*) indicating the degree to which they felt that the client was emotionally expressive, comfortable expressing affection toward men, emotionally and physically tough, desirous of status and success at work, comfortable with his feminine side, and "traditional" in terms of his attitudes and behaviors. A MANOVA was conducted, with the traditionalism of the male videotape as the independent variable and the six aforementioned questions as the dependent variables. The MANOVA was statistically significant, $F(6, 45) = 66.70$, $p < .05$; the participants perceived the traditional client as more traditional than the nontraditional client across all six questions. Furthermore, participants' average plausibility and believability ratings, using a Likert-type scale (1 = *strongly disagree,* 7 = *strongly agree*), were 5.5 and 5.2, respectively, for the nontraditional

male client condition and 6.1 and 5.7, respectively, for the traditional male client condition, indicating that participants generally perceived both of the videotaped male client vignettes to be at least moderately plausible and believable.

Measures

Male Role Norms Scale (MRNS; Thompson & Pleck, 1986). The MRNS measures masculine ideology. It contains 26 items rated on a 7-point Likert-type scale (1 = *strongly disagree,* 7 = *strongly agree*); higher scores reflect more traditional male gender role attitudes (Thompson & Pleck, 1986). The MRNS contains three subscales: Status (e.g., "Success in his work has to be a man's central goal in life"), Toughness (e.g., "When a man is feeling a little pain he should not let it show very much"), and Anti-femininity (e.g., "It is a bit embarrassing for a man to have a job that is usually filled by a woman"; Thompson & Pleck, 1986). The MRNS was modified to help obscure the purpose of the study by adding 26 distracter items.

We used the total MRNS score, which displays excellent reliability (alpha = .86; Thompson & Pleck, 1986). The internal consistency (coefficient alpha) of MRNS scores in this study was .82. MRNS scores are directly related to homophobia, adversarial views of sexual relationships, and traditional careers (Jome & Tokar, 1998; Sinn, 1997). The MRNS has displayed construct validity with both men and women (Stark, 1991; Thompson, 1990).

Therapist Appraisal Questionnaire (TAQ). The TAQ was used to assess counselors' primary appraisals. Fauth et al. (1999) created the TAQ by modifying a scale originally developed by Cooley and Klingler (1989) for studying academic stress. The TAQ consists of 16 items rated on a 5-point Likert-type scale (0 = *not at all,* 4 = *a great deal*) and three scales: Challenge (i.e., how exhilarated, hopeful, pleased, eager, happy, energetic, and excited counselors were when counseling their client), Threat (i.e., how worried, fearful, anxious, and confident counselors were when counseling their client), and Harm (i.e., how angry, sad, guilty, disappointed, and disgusted counselors were when counseling their client; Cooley & Klingler, 1989). We combined the Threat and Harm scales, which are moderately to highly correlated (Fauth et al., 1999), to create a Negative Stress scale.

The TAQ scales possess excellent internal consistency, with alpha coefficients ranging from .71 to .90 (Cooley & Klingler, 1989; Fauth et al., 1999). In this study, internal consistency (coefficient alpha) for the Challenge and Negative Stress scales were .86 and .74, respectively. The TAQ has also displayed construct validity. For example, Challenge and Threat scores were associated, as predicted by the transactional theory of stress, with self-efficacy and prognosis scores in a prior countertransference investigation (Fauth et al., 1999).

The TAQ assesses counselors' primary appraisals indirectly via the emotions associated with harm/loss, threat, and challenge appraisals. Direct measurement would require

counselors to have insight into the degree to which the client presented a threat, harm, or challenge to their personally meaningful goals or needs. Yet, it is probable that counselors (like everyone else) often lack insight into the factors contributing to their subjective experiences even though they can accurately report those subjective experiences (Nisbett & Wilson, 1977).

Counseling Self-Estimate Inventory (COSE; Larson et al., 1992). We used the Awareness of Values (AV) subscale from the COSE to assess participants' secondary appraisals (i.e., self-efficacy). The AV contains four items rated on a 6-point Likert-type scale (1 = *strongly disagree,* 6 = *strongly agree*). The AV subscale assesses counselors' perceived efficacy in successfully managing their self-awareness and value conflicts with clients (Larson et al., 1992). The COSE displays excellent internal consistency (alpha = .86; Fauth et al., 1999), although the internal consistency of the AV tends to be lower than that of the total score. In this study, the internal consistency (coefficient alpha) of the AV was .60. Countertransference research revealed that the COSE scales were directly related to challenge appraisals and performance expectations but inversely related to threat appraisals (Fauth et al., 1999).

Avoidance Index (Bandura et al., 1960). The Avoidance Index assesses counselors' behavioral countertransference manifestations by categorizing counselors' verbal responses into approach or avoidance categories (Bandura et al., 1960). Approach responses encourage the client to continue exploring the current therapy content. In contrast, avoidance responses inhibit or divert the client from further exploration. Approach responses are approval (i.e., agreeing with the client), exploration (i.e., asking for elaboration), instigation (i.e., redirecting toward the topic), reflection (i.e., paraphrasing), and labeling (i.e., interpretation of feelings). Avoidance responses are disapproval (i.e., being critical of the client), topical transition (i.e., changing the topic), silence (i.e., no response), ignoring (i.e., not responding to feelings), and mislabeling (i.e., misinterpreting feelings).

Three student raters, who were not aware of the study's hypotheses, were trained to implement the approach–avoidance categories using commercially available training tapes. The raters classified therapist responses into 1 of 11 categories (i.e., 5 approach categories, 5 avoidance categories, and 1 unclassifiable category), using review and discussion to aid in the training process. Training consisted of three 3-hour group training sessions and was discontinued when interrater agreement reached .80. The interrater reliabilities *(r)* between the avoidance indices for the three pairs of raters were .77, .76, and .74. Interrater agreement scores regarding whether each therapist response was primarily avoidance, approach, or unclassifiable were .94, .90, and .91 for the three pairs of raters.

Raters then classified each therapist response. For each therapist, we calculated a cumulative ratio of the number of avoidance responses to the number of approach and avoidance responses. The Avoidance Index has displayed strong interrater reliability, interrater agreement, and construct validity in countertransference investigations (Bandura et al., 1960; Gelso et al., 1995; Hayes & Gelso, 1993; Latts & Gelso, 1995; Rosenberger & Hayes, 2002). In fact, the Avoidance Index is the only existing measure of countertransference for which strong evidence of construct validity exists.

GAF (APA, 1994). Counselors' GAF ratings of the client were used to assess counselors' cognitive countertransference manifestations. The GAF is used by clinicians to rate clients' psychological, social, and occupational functioning. GAF scores for these ratings range from 0 to 100, with higher scores reflecting more adaptive functioning (APA, 1994). GAF reliability varies from marginal to excellent (Jones, Thornicroft, Coffey, & Dunn, 1995), with test–retest (intraclass) reliability ranging from .47 to .82 (J. B. W. Williams et al., 1992). Although the reliability of the GAF has been questioned, the scale is clinically important because of its widespread use as an aid in treatment referral and planning. One reason that GAF scores may tend to be unreliable is the subjective nature of the scale, which makes it a potentially useful scale to measure clinical biases and countertransference reactions. In fact, GAF scores have been successfully used to study counselors' clinical biases toward male clients (Wisch & Mahalik, 1999). Furthermore, GAF scores do seem to possess construct validity in that they have successfully differentiated between diverse diagnostic groups (Goldman, Skodol, & Lave, 1992). GAF scores have also displayed convergent validity with other measures of occupational and social functioning (Goldman et al., 1992; Jones et al., 1995).

Hesitance. Counselors' hesitance in responding to the client was used as an experimental measure of nonverbal countertransference behavior. The team of approach–avoidance raters recorded the amount of time (in seconds) that elapsed between the end of each client segment and the beginning of each therapist response. Although this measure has not been used previously in countertransference research, response latency has been used as an indicator of anxiety and stress in other psychological research (e.g., Foa, Feske, Murdock, Kozak, & McCarthy, 1991). Pauses of 4 seconds or longer are typically coded as avoidance responses when coding counselors' responses for the Avoidance Index. Because we thought that counselors' hesitance might indeed capture a nonverbal aspect of counselors' countertransference reactions to clients but were wary of the arbitrary nature of the 4-second cutoff mark, we decided to extract the code from the Avoidance Index and use it as a separate continuous variable. Thus, we included this variable as an exploratory measure in this study with the purpose of ascertaining its relevance to countertransference research.

Procedure

Counselors first completed an informed consent form along with the MRNS. About 1 week later, counselors met one of

the experimenters in a counseling room in the department's training clinic or the counseling center. Upon their arrival, counselors received instructions regarding viewing and responding to the videotaped client vignette. They were asked to assume the mental set of a third therapy session. They were instructed to make therapeutic interventions "as if" they were speaking directly to an actual client. Counselors were informed that a "Please Respond" screen would appear after each of the five client segments; they were instructed to pause the tape at that time and make their response.

Counselors were randomly assigned by gender into either the traditional or the nontraditional male client condition. Counselors viewed the videotape and responded into the microphone of a continuously operating tape recorder. After completing the videotape, counselors completed the TAQ, the AV, and the GAF, which were counterbalanced to control for order effects.

◼ Results

For both research questions, we used a two-tiered analysis strategy consisting of an omnibus test followed by more targeted analyses, thereby decreasing family-wise error by eliminating unnecessary analyses.

Hypothesis 1

Counselors' stress appraisals were expected to predict their countertransference manifestations (see Table 1 for bivariate correlations). Thus, a multivariate multiple regression analysis was conducted with counselors' avoidance, GAF, and hesitance scores as the criterion variables and Challenge, Negative Stress, and AV scores as the predictor variables. Results indicated that the predictor variables were statistically significantly related to the criterion variables, $F(3, 56) = 26.36$, $p \le .01$.

Because the MANOVA revealed significant results, we conducted a series of follow-up hierarchical regression analyses to further specify the relationships between the stress appraisal and the countertransference manifestation variables.

TABLE 1

Correlations Among the Stress Appraisal and Countertransference Manifestation Variables

Variable	1	2	3	4	5	6
1. Challenge	—	−.34**	−.05	−.15	.23*	−.18
2. Negative Stress		—	−.27**	−.04	.11	.34**
3. AV			—	−.30**	−.06	−.31**
4. Avoidance				—	−.15	.14
5. GAF					—	−.19
6. Hesitance						—

Note. AV = Awareness of Values subscale. GAF = Global Assessment of Functioning Scale.
*$p \le .05$, one-tailed. **$p \le .01$, one-tailed. ***$p \le .001$, one-tailed.

In these analyses, counselors' avoidance, GAF, and hesitance scores were regressed onto counselors' Challenge and Negative Stress scores (i.e., primary appraisals) in Step 1 and AV scores (i.e., secondary appraisals) in Step 2 (see Table 2). We entered these variables in separate steps to assess whether therapists' secondary appraisals would add to the prediction of countertransference behavior achieved by using therapists' primary appraisals alone (in past transactional research in nontherapy settings, primary appraisals have been better predictors than secondary appraisals).

Counselors' Challenge and Negative Stress scores accounted for 3% of the variance in counselors' avoidance, which was not statistically significant ($R = .18$, $F = 1.02$, $p > .05$). AV scores accounted for an additional 13% of the variance in avoidance scores, which was statistically significant ($F = 8.88$, $p \le .05$). Counselors' perceived efficacy in managing their awareness of emotional reactions to and value conflicts with the client significantly added to the prediction of their avoidance behavior beyond the contribution of primary appraisals. The full regression model accounted for 16% of the variance in counselors' avoidance behavior, which was also statistically significant ($R = .40$, $F = 3.72$, $p \le .05$).

Counselors' Challenge and Negative Stress scores accounted for 9% of the variance in counselors' GAF scores, which was statistically significant ($R = .31$, $F = 3.06$, $p \le .05$).

TABLE 2

Hierarchical Multiple Regression Results for Avoidance, GAF, and Hesitance Scores Regressed Onto Challenge, Negative Stress, and AV Scores

Criterion Variable	R	ΔR^2	β
Avoidance			
Step 1: *df* = 2, 61	.18	.03	
Challenge			−.19
Negative Stress			−.11
Step 2: *df* = 3, 60	.40*	.13**	
Challenge			−.24
Negative Stress			−.23
AV			−.37**
GAF			
Step 1: *df* = 2, 61	.31*	.09*	
Challenge			.30*
Negative Stress			.21
Step 2: *df* = 3, 60	.31	.00	
Challenge			.31*
Negative Stress			.22
AV			.01
Hesitance			
Step 1: *df* = 2, 61	.35*	.12*	
Challenge			−.07
Negative Stress			.32*
Step 2: *df* = 3, 60	.42**	.05*	
Challenge			−.11
Negative Stress			.24
AV			−.25*

Note. AV = Awareness of Values subscale. GAF = Global Assessment of Functioning Scale.
*$p \le .05$. **$p \le .01$.

Beta weights indicated that Challenge scores were unique predictors of counselors' GAF scores; specifically, as counselors experienced more positive reactions toward the client, their GAF scores increased. AV scores did not account for any additional variance in GAF scores ($F = 0.01$, $p > .05$). The full regression model after Step 2 accounted for 9% of the variance in counselors' GAF scores, which was not statistically significant ($R = .31$, $F = 2.40$, $p > .05$).

Counselors' Challenge and Negative Stress scores accounted for 12% of the variance in counselors' hesitance scores, which was statistically significant ($R = .35$, $F = 4.30$, $p \leq .05$). Beta weights indicated that Negative Stress scores uniquely predicted counselors' hesitance. As counselors appraised the client as more threatening and harmful, their hesitance increased. AV scores accounted for an additional 5% of variance in hesitance scores, which was statistically significant ($F = 3.95$, $p \leq .05$). As counselors felt less able to manage their reactions toward the client, they became more hesitant in responding to him. The full regression model after Step 2 accounted for 18% of the variance in counselors' hesitance, which was statistically significant ($R = .42$, $F = 4.31$, $p \leq .01$).

Hypothesis 2

The second hypothesis predicted that counselors' stress appraisals should mediate any interaction between counselors' male gender role attitudes and client male gender role conformity in predicting counselors' countertransference manifestations. To establish such a mediation effect, we first needed to assess whether counselor's male gender role attitudes and client male gender role conformity did in fact interact to predict therapists' stress appraisals and countertransference manifestations. We used hierarchical multivariate multiple regression (Finn & Bock, 1988) as an omnibus test of the aforementioned interaction effect. First, the main effect (i.e., MRNS scores and client condition) and interaction variables were centered to reduce redundancy (Aiken & West, 1991). Next, the main effect variables were isolated by entering them in Step 1 (Baron & Kenny, 1986), followed by the interaction term in Step 2. The criterion variables were counselors' Challenge, Negative Stress, AV, avoidance, GAF, and hesitance scores. The statistical significance of the change in variance accounted for after each step, as well as the statistical significance of the full regression model, was examined to evaluate the interaction hypothesis. Neither Step 1, $F(12, 108) = 1.12$, $p > .05$, Step 2, $F(18, 150) = 0.90$, $p > .05$), nor the change in variance accounted for between Step 1 and Step 2, $F(6, 53) = .37$, $p > .05$, was statistically significant. Because no interaction effect emerged, it was not possible to fully test the mediation hypothesis.

Discussion

The transactional conceptualization of countertransference received some support in this study. We describe these find-

ings in the following section and discuss them in light of potential future research and the methodological limitations of the investigation. We conclude by presenting potential clinical implications of the transactional view of countertransference.

Relationship Between Stress Appraisals and Countertransference Manifestations

We assessed the relationship between counselors' stress appraisals and their countertransference manifestations using multiple regression analyses. Despite the fact that only 5% of counselors' responses were classified as avoidant (thereby making prediction of this variable quite difficult), counselors' stress appraisals predicted more than 16% of the variance of this variable. This effect is robust when compared with that of other countertransference investigations (Gelso et al., 1995; Hayes & Gelso, 1993) and is potentially clinically meaningful because previous research has indicated that the kinds of behaviors categorized as avoidant have been inversely linked with counseling outcome (Henry, Schact, & Strupp, 1986). Counselors' secondary appraisals were particularly potent predictors of their avoidance behavior. Specifically, counselors who felt more efficacious in managing their feelings and value conflicts with the client tended to avoid him less. Thus, helping counselors to feel more confident in successfully managing their emotional reactions and value conflicts during the session may also be helpful in reducing their avoidance behavior.

Primary stress appraisals, especially challenge appraisals, were unique predictors of counselors' GAF scores. Specifically, as counselors perceived the counseling situation as more of an opportunity for personal gain, they rated the client's functioning more positively. Clinically, if counselors' GAF scores are unduly influenced by their challenge appraisals, it could lead to their making inappropriate treatment referrals and decisions. In this era of rapid assessment and brief treatment, the influence of counselors' primary stress appraisals on their diagnostic formulations could be especially damaging to client care. Thus, the role of counselors' primary appraisals in relation to clinical biases and treatment decisions warrants further study.

Primary and secondary stress appraisals predicted counselors' hesitance. As counselors felt more threatened and harmed by the client and less able to manage their reactions, their hesitance in responding to him increased. Thus, counselors' stress appraisals were associated with decreased spontaneity, which *could* damage the therapeutic alliance. For instance, clients might interpret therapist hesitance as an attempt by the therapist to conceal her or his true reactions. Of course, hesitance might also be viewed in a more positive light, as a thoughtful way for therapists to manage their personal reactions. Nonetheless, some research evidence suggests that clients perceive their counselors as less helpful when counselors report being most aware of their own personal reactions (E. N. Williams, 1999). Furthermore, the

fact that hesitance covaried with counselors' negative stress appraisals suggests that the measure of hesitance, at least in part, tapped into a nonverbal manifestation of counselors' negative reactions (and did not simply measure some stable counselor characteristic such as a slow speech pattern or thoughtfulness).

In sum, these results support the notion that counselors' stress appraisals are linked with their countertransference manifestations, lending some support to the transactional reconceptualization of countertransference proposed in this article (Hayes et al., 1998). Moreover, as one might expect, more positive (i.e., challenge) stress appraisals seem to be associated with more positive feelings and actions toward clients (i.e., more positive diagnostic evaluations of clients), whereas more negative stress appraisals (i.e., feeling threatened and or harmed by the client and unable to manage those reactions) seem to be associated with more negative feelings and actions toward clients (i.e., increased hesitance, increased avoidance behaviors).

Interaction Effect

We did not find that counselors' male gender role attitudes and male clients' gender role conformity interacted to predict counselors' stress appraisals and countertransference manifestations. The lack of an interaction was somewhat puzzling in light of previous research. Perhaps counselors' consistently nontraditional male gender role attitudes were responsible for the lack of an interaction effect. This explanation seems plausible, because only one therapist endorsed traditional male gender role attitudes (based on the midpoint of the MRNS). The restricted range of counselors' male gender role attitudes likely attenuated any potential interaction effect.

Alternatively, the interaction hypothesis itself may have been faulty. However, counselors' male gender role attitudes have previously been implicated in their reactions to male clients (Wisch & Mahalik, 1999). Thus, it seems possible that a mismatch between counselors' and clients' male gender role conformity could trigger counselors' stress appraisals and countertransference manifestations. Most likely, countertransference occurs when this mismatch is most extreme. Only future research, however, can assess the veracity of this assumption.

Methodological Limitations

Three methodological limitations constrain the extent to which we can confidently generalize about these findings. The most serious limitation pertains to the analogue design of the study and the attendant concerns regarding external validity. In addition, we investigated counselors' countertransference only with White male clients. Thus, caution should be exercised in generalizing these results to White female clients and male and female clients of color. Given the importance of multicultural counseling, counselors'

countertransference in response to women and men of color warrants increased research attention. Finally, it should be noted that counselors' stress appraisals were assessed solely with self-report instruments. Future research would profit from using multiple measures (including coding of qualitative data) to assess counselors' stress appraisals.

Future Research

Clearly, the results of this study suggest that further investigation of the transactional model of countertransference is warranted. It is also important to remember that the results of this study were mixed and, as such, are far from conclusive about the applicability of the model. Indeed, much future research will be needed to establish the transactional model as an appropriate heuristic for countertransference research. To further establish the transactional theory of countertransference, future lab research would profit from more salient and realistic client conditions, such as role-play scenarios with experimenter confederates, which would retain considerable internal validity while enhancing external validity. Furthermore, more countertransference research should be conducted in the field to examine the relationships between counselors' stress appraisals, counselors' countertransference manifestations, therapy process, and therapy outcome. Finally, the success of the hesitance measure suggests that researchers should develop measures of nonverbal countertransference manifestation. Such measures could be coupled with behavioral assessment techniques such as time sampling (e.g., random sampling and interval recording) to increase data collection efficiency.

Clinical Implications

Should the transactional theory of countertransference prove to be a useful heuristic, counselors could systematically assess their stress appraisals, thereby heightening their countertransference awareness. Quantitatively inclined counselors could use the TAQ and AV to monitor their stress appraisals; qualitatively inclined counselors could examine their process notes for positive and negative appraisal themes. In possession of new insight into their personal reactions to clients, counselors could adjust their therapeutic approach and, when necessary, seek appropriate supervision and consultation to help them successfully manage their reactions.

The transactional theory could also guide the exploration of countertransference issues in supervision. For emotionally attuned trainees, supervisors might initially focus on trainees' stress appraisals. For example, trainees with excessively negative stress appraisals (negative feelings and low perceived efficacy in managing those feelings) might withdraw from clients. Subsequent supervision might profitably explore the countertransference triggers and origins at hand. For less emotionally attuned trainees, supervision might initially explore trainees' actual clinical behavior. Once coun-

tertransference behavior has been identified (e.g., avoidance or hesitance), supervisors could facilitate exploration into trainees' stress appraisals, countertransference triggers, and countertransference origins. For example, if a trainee withdraws from a client, supervision might explore the personally threatening or harmful aspects of the therapeutic situation. Whatever the starting point, supervision guided by the transactional theory might help clarify the particular countertransference dynamics at hand and enhance trainees' ability to identify and resolve future countertransference issues.

According to the transactional theory, it is both normal and inevitable that counselors will appraise (often automatically and unconsciously) the therapy environment in terms of their own well-being. Thus, perhaps the most powerful potential contribution of the transactional theory is that it reframes countertransference as inevitable, functional, and normal, thus counteracting the widespread view that countertransference is pathological and can or should be avoided. It is hoped that this normalization will empower counselors to acknowledge, understand, and use their countertransference reactions to their own and to their clients' benefit.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 2,* 122–147.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bandura, A., Lipsher, D. H., & Miller, P. E. (1960). Psychotherapists' approach-avoidance reactions to patients' expressions of hostility. *Journal of Consulting Psychology, 24,* 1–8.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Cooley, E. J., & Klingler, C. R. (1989). Academic attributions and coping with tests. *Journal of Social and Clinical Psychology, 8,* 359–367.

Cutler, R. L. (1958). Countertransference effects in counseling. *Journal of Consulting Psychology, 22,* 349–356.

Fauth, J., Hayes, J. A., Park, J., & Friedman, S. (1999, August). *Presession countertransference: Counselors' appraisals, self-efficacy, and prognosis.* Poster session presented at the annual meeting of the American Psychological Association, Boston.

Finn, J., & Bock, R. D. (1988). *Multivariance PC* [Computer software]. Mooresville, IN: Scientific Software.

Fitzgerald, L. F., & Cherpas, C. C. (1985). On the reciprocal relationship between gender and occupation: Rethinking the assumptions concerning masculine career development. *Journal of Vocational Behavior, 27,* 109–122.

Foa, E. B., Feske, U., Murdock, T. B., Kozak, M. J., & McCarthy, P. R. (1991). Processing of threat-related information in rape victims. *Journal of Abnormal Psychology, 100,* 156–162.

Folkman, S. (1984). Personal control and stress and coping processes: A theoretical analysis. *Journal of Personality and Social Psychology, 46,* 839–852.

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: A study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology, 48,* 150–170.

Freud, S. (1959). Future prospects of psychoanalytic psychotherapy. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 20, pp. 87–172). London: Hogarth. (Original work published 1910)

Fromm-Reichmann, F. (1950). *Principles of intensive psychotherapy.* Chicago: University of Chicago.

Gelso, C. J., & Carter, J. A. (1985). The relationship in counseling and psychotherapy: Components, consequences, and theoretical antecedents. *The Counseling Psychologist, 13,* 155–243.

Gelso, C. J., & Carter, J. A. (1994). Components of the psychotherapy relationship: Their interaction and unfolding during treatment. *Journal of Counseling Psychology, 41,* 296–306.

Gelso, C. J., Fassinger, R. E., Gomez, M. J., & Latts, M. G. (1995). Countertransference reactions to lesbian clients: The role of homophobia, counselor gender, and countertransference management. *Journal of Counseling Psychology, 42,* 356–364.

Gelso, C. J., & Hayes, J. A. (1998). *The counseling relationship: Theory, research, and practice.* New York: Wiley.

Goldman, H. H., Skodol, A. E., & Lave, T. R. (1992). Revising Axis V for *DSM-IV:* A review of measures of social functioning. *American Journal of Psychiatry, 149,* 1148–1156.

Good, G. E., Robertson, J. M., O'Neil, J. M., Fitzgerald, L. F., Stevens, M., DeBord, K. A., et al. (1995). Male gender role conflict: Psychometric issues and relations to psychological distress. *Journal of Counseling Psychology, 42,* 3–10.

Hayes, J. A. (1995). Countertransference in group counseling: Waking a sleeping dog. *International Journal of Group Counseling, 45,* 521–535.

Hayes, J. A., & Gelso, C. J. (1991). Effects of therapist-trainees' anxiety and empathy on countertransference behavior. *Journal of Clinical Psychology, 47,* 284–290.

Hayes, J. A., & Gelso, C. J. (1993). Counselors' discomfort with gay and HIV-infected clients. *Journal of Counseling Psychology, 40,* 86–93.

Hayes, J. A., McCracken, J. E., McClanahan, M. K., Hill, C. E., Harp, J. S., & Carozzoni, P. (1998). Therapist perspectives on countertransference: Qualitative data in search of a theory. *Journal of Counseling Psychology, 45,* 468–482.

Henry, W. P., Schact, T. E., & Strupp, H. H. (1986). Structural analysis of social behavior: Application to a study of interpersonal process in differential psychotherapeutic outcome. *Journal of Consulting and Clinical Psychology, 54,* 27–31.

Jome, L. M., & Tokar, D. M. (1998). Dimensions of masculinity and major choice traditionality. *Journal of Vocational Behavior, 52,* 120–134.

Jones, S. H., Thornicroft, G., Coffey, M., & Dunn, G. (1995). A brief mental health outcome scale: Reliability and validity of the Global Assessment of Functioning (GAF). *British Journal of Psychiatry, 166,* 654–659.

Kiesler, D. J. (1982). Interpersonal theory of personality and psychotherapy. In J. C. Anchin & D. J. Kiesler (Eds.), *Handbook of interpersonal psychotherapy* (pp. 3–24). New York: Pergamon Press.

Larson, L. M., Suzuki, L. A., Gillespie, K. N., Potenza, M. T., Bechtel, M. A., & Toulouse, A. L. (1992). Development and validation of the Counseling Self-Estimate Inventory. *Journal of Counseling Psychology, 39,* 105–120.

Latts, M. G., & Gelso, C. J. (1995). Countertransference behavior and management with survivors of sexual assault. *Journal of Counseling Psychology, 32,* 405–415.

Lazarus, R. S. (1993). Why we should think of stress as a subset of emotion. In L. Goldberger & S. Breznitz (Eds.), *Handbook of stress: Theoretical and clinical aspects* (pp. 21–39). New York: Free Press.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York: Springer.

Levenson, H. (1995). *Time-limited dynamic counseling.* New York: Basic Books.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

O'Neil, J. M. (1980). Male sex role conflicts, sexism, and masculinity: Psychological implications for men, women, and the counseling psychologist. *The Counseling Psychologist, 9,* 61–80.

Robertson, J. M., & Fitzgerald, L. F. (1990). The (mis)treatment of men: Effects of client gender role and life-style on diagnosis and attribution of pathology. *Journal of Counseling Psychology, 37,* 3–9.

Rosenberger, E. W., & Hayes, J. A. (2002). Causes, consequences, and management of countertransference: A case study. *Journal of Counseling Psychology, 49,* 221–232.

Sherman, M. (1965). Peripheral cues and the invisible countertransference. *American Journal of Counseling, 19,* 280–292.

Singer, B. A., & Luborsky, L. (1977). Countertransference: The status of clinical versus quantitative research. In A. S. Gurman & A. M. Razdin (Eds.), *Effective counseling: Handbook of research* (pp. 433–451). New York: Pergamon Press.

Sinn, J. S. (1997). The predictive and discriminant validity of masculinity ideology. *Journal of Personality, 31,* 117–135.

Stark, L. P. (1991). Traditional gender role beliefs and individual outcomes: An exploratory analysis. *Sex Roles, 24,* 639–650.

Stein, N. L., Folkman, S., Trabasso, T., & Richards, T. A. (1997). Appraisal and goal processes as predictors of well-being in bereaved caregivers. *Journal of Personality and Social Psychology, 72,* 872–884.

Strupp, H. H. (1980). Success and failure in time-limited counseling: A systematic comparison of two cases: Comparison 2. *Archives of General Psychiatry, 37,* 708–716.

Thompson, E. H. (1990). Courtship violence and the male role. *Men's Studies Review, 7,* 4–13.

Thompson, E. H., & Pleck, J. H. (1986). The structure of male role norms. *American Behavioral Scientist, 29,* 531–543.

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat and challenge appraisals. *Journal of Personality and Social Psychology, 73,* 63–72.

Williams, E. N. (1999, November). *From anxiety to boredom: Managing self-awareness and therapist reactions in therapy.* Paper presented at the annual meeting of the Mid-Atlantic Chapter of the Society for Counseling Research, College Park, MD.

Williams, J. B. W., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., et al. (1992). The structured clinical interview for *DSM-III-R* (SCID): Multisite test-retest reliability. *Archives of General Psychiatry, 49,* 630–637.

Wisch, A. F., & Mahalik, J. R. (1999). Male counselors' clinical bias: Influence of client gender roles and therapist gender role conflict. *Journal of Counseling Psychology, 46,* 51–60.

Yulis, S., & Kiesler, D. J. (1968). Countertransference response as a function of therapist anxiety and content of patient talk. *Journal of Consulting and Clinical Psychology, 32,* 413–419.

TRANSFERENCE AND COUNTERTRANSFERENCE IN COGNITIVE BEHAVIORAL THERAPY

Jan Prasko[a,b,c,d]*, Tomas Diveky[a,b] , Ales Grambal[a,b], Dana Kamaradova[a,b], Petr Mozny[e], Zuzana Sigmundova[a,b], Milos Slepecky[f], Jana Vyskocilova[c]

[a] *Department of Psychiatry, University Hospital Olomouc, I. P. Pavlova 6, 775 20 Olomouc, Czech Republic*
[b] *Psychiatric Clinic, Faculty of Medicine and Dentistry, University Palacky Olomouc*
[c] *Prague Psychiatric Centre, Ustavni 91, Prague 8*
[d] *Centre of Neuropsychiatric Studies, Ustavni 91, Prague 8*
[e] *Mental Hospital Kromeriz*
[f] *Psychagogia s.r.o, Liptovsky Mikulas, Slovakia*
E-mail: prasko@fnol.cz;

Background. Both patients and psychotherapists can experience strong emotional reactions towards each other in what are termed transference and countertransference within therapy. In the first part of this review, we discuss transference issues. Although not usually part of the obvious language of cognitive behavioral therapy (CBT), examination of the cognitions related to the therapist, is an integral part of CBT, especially in working with difficult patients.

In the second part, we cover counter-transference issues. We describe schematic issues that give rise to therapist counter-transference and explain how this interacts in different types of patient therapist encounter. We also examine ways in which the therapist can use CBT to help him/her modify the countertransference and, in the process, assist the patient.

Methods. PUBMED data base was searched for articles using the key words "therapeutic relations", "transference", "countertransference", "cognitive behavioral therapy", "cognitive therapy", "schema therapy", "dialectical behavioral therapy". The search was repeated by changing the key word. No language or time constraints were applied. The lists of references of articles detected by this computer data base search were examined manually to find additional articles. We also used the original texts of A. T. Beck, J. Beck, M. Linehan, R. Leahy, J. Young and others. Basically this is a review with conclusions about how therapists can manage transference issues.

Results. Transference. The therapist should pay attention to negative or positive reactions towards him/ her but should not deliberately provoke or ignore them. He/she should be vigilant for signs of strong negative emotions, such as a disappointment, anger, and frustration experienced in the therapeutic relationship by the patient. Similarly he/she should be alert to exaggerated positive emotions such as love, excessive idealization, praise or attempts to divert the attention of therapy onto the therapist. These reactions open space for understanding the patient's past and actual relations outside the therapy. Countertransference. The therapist should be aware of countertransference schemas as they apply to him/her. He/she should monitor his/her own feelings that indicate countertransference. Further, the assistance of and discussion with supervisors and colleagues is useful in regard to countertransference even in experienced therapists. Countertransference can be used as an open window into the interpersonal relations of the patient.

Conclusions. Both the literature and our experience underscore the importance of careful and open examination of both transference and counter-transference issues in CBT and their necessary incorporation in the complete management of all patients undergoing CBT.

INTRODUCTION

Transference is a phenomenon in psychoanalysis characterized by unconscious redirection of feelings from one person to another, in the case of psychotherapy from patient to therapist. Countertransference refers to transference of feelings from therapist to patient. Both phenomena are based on the human ability to recognise the outside world. Whenever a person perceives something, in order to recognise it, he/she needs to compare the present perception with past experience. Memories, especially pictorial and emotionally laden ones, can be triggered by something in the present which evokes the past and the person will react in a similar way as they did then[1].

Transference and countertransference feelings/ reactions are a valuable source of information about a patient's (and therapist's) inner world. Some psychotherapists have had or have the perception that transference is not examined in cognitive behavioral therapy (CBT). They believe that the general technical difference between psychodynamic psychotherapy and cognitive behavioral therapy (CBT) is emphasis of the former on transference. This was one of the major misconceptions of CBT that was identified by various experts[2-5].

Psychodynamic therapists view transference as a powerful tool in understanding the patient and eventually effecting change. They believe that cognitive behavioral therapists eschew transference as an intervention that would distract the patient from outside relationships and risk therapeutic rupture[6]. Although the interpretation of transference is not a central tool of cognitive therapy, automatic thoughts and feelings related to interactions with the therapist are very much within the scope of exploration and may provide valuable opportunities for testing and modifying dysfunctional automatic thoughts[2,7]. A good therapeutic relationship is an important issue for the effective treatment in cognitive behavioral therapy. Cognitive behavioral therapists generally aim to establish an open collaborative relationship at the start of therapy and then to work directly towards them without paying too much attention to interpersonal issues. Clinical competence, conviction, and consistency seem to predict a more successful psychotherapeutic outcome[8-10]. However, when working with difficult patients (e.g. patients with personality disorder, hypochondriasis etc.) psychotherapy is rarely strightforward. The dysfunctional schemas, beliefs and assumptions that bias the patient's perceptions of others are likely to bias their perception of the therapist. The dysfunctional interpersonal behaviour strategies, manifest in the patient-therapist relationship. If they are not addressed effectively, interpersonal difficulties arising in the patient-therapist relationship can disrupt the therapy. However, these difficulties also provide the therapist with an opportunity to directly observe an intervention rather than having to rely on the patient's report of interpersonal problems occuring outside the sessions[11]. Therefore transference and countertransference feelings/reactions are a valuable source of information about a patient's (and therapist's) inner world.

TRANSFERENCE FROM THE COGNITIVE BEHAVIORAL POINT OF VIEW

Although the word "transference" is not part of the jargon of cognitive behavioral therapy, examination of the cognitions related to the therapist apropos past significant relationships is an integral part of assessment and treatment within CBT[12,13]. The patient's emotional reaction to the dynamics of therapy and therapist are important especially in working with difficult patients. Always alert but not provoking, the therapist should be ready to explore these reactions for more information about the patient's system of thoughts and beliefs. If not explored, possible distorted interpretations persist and may interfere with the collaboration. If exposed, they often provide rich material for understanding the meanings and beliefs behind the patient's idiosyncratic or repetitious reactions. Some forms of therapy depend on the use of transference. Simply stated, transference occurs when the patient superimposes prior experiences on the therapist[14]. The patient may perceive something about the therapists personality, style, demeanor, or appearance that may remind him or her of a significant person in the patient's past, such as a parent; hence, the patient may begin to respond to the therapist as the patient would to that parent. Effective transference is facilitated by the therapist persistently withholding self-disclosure, presenting as a virtually unbiased blank slate (tabula rasa) wherein the patient is free to superimpose (transfer) his or her feelings for a significant person in his or her own life (a parent, spouse, sibling, peer, etc.) onto the neutral therapist[15]. This is not the aim of CBT. Self-disclosure, warm and empathetic atmosphere, collaborative relation and stress on the self-efficacy of the client may pollute this transference potential[14].

There are many telltale signs of transference. These are the same signs that suggest the presence of automatic thoughts during the session. E.g. there may be a sudden change in the patient's nonverbal behavior: sudden change in expression, abruptly switching to a new topic, stammer, block, pauses in the middle of a train of statements, slumping posture, clenching fists, kicking, tapping foot and so on. One of the most revealing signs is a shift in the patient's gaze, especially if he/she has had a thought but prefers not to reveal it[16]. When asked, the patient may say, "It is not important." The therapist should press the patient nonetheless, gently, as it might be important.

The therapist should pay attention to any negative or positive reactions to him/her that arise but should not deliberately provoke or ignore them. He/she should be vigilant for signs of disappointment, anger, and frustration experienced by the patient in the therapeutic relationship. Similarly the therapist should be alert to excessive idealization, praise or attempts to divert the attention of therapy onto the therapist. These reactions open windows into the patient's past and actual relations outside the therapy. The therapist would be unable to view the meanings or beliefs beyond these windows if the arousal of their own affective responses is viewed as a distraction to be controlled, avoided, or suppressed. Hoffart et al.[17] examined whether therapists' emotional reactions to their patients mediate the effect of personality disorders and interpersonal problem behaviours on the outcome of treatment with a focus on an Axis I disorder and, whether the therapists' reactions mediate the effect of personality disorders on the course of interpersonal problems. Therapists completed a checklist of emotional reactions to individual patients after the end of residential cognitive or guided mastery therapy for 46 inpatients with panic disorder with agoraphobia. The severity of DSM-III-R personality disorder was related to the therapists' insecurity feelings but not to interest or anger. A higher level of therapist insecurity feelings was related to less reduction in self-reported agoraphobic avoidance during treatment, whereas the therapists' emotions were unrelated to symptomatic course after treatment. Therapists' insecurity feelings appeared partly to mediate the relationship between patient severity of personality disorder and persistence of patients' interpersonal dominance and nurturance problems.

Rossberg et al.[18] studied the relationship between patients' self-reported personality characteristics, treatment outcome and therapists' countertransference reactions. Eleven therapists filled in the Feeling Word Checklist 58 (FWC-58) for each patient admitted to a day treatment program. The patients completed the Circumplex of Interpersonal Problems (CIP) at admission and discharge. At the end of treatment, the outcome measures were assessed. At the start of treatment, therapists reported fewer feelings of rejection and being on guard in response to patients who reported high avoidant, exploitable, overly nurturing and intrusive CIP subscale traits. At the end of the treatment, the CIP subscales of being domineering, vindictive and cold correlated with fewer positive and more negative countertransference feelings. The study revealed a strong relationship between improvement and countertransference feelings. This study confirms clinical narratives of relationships between the therapists' countertransference reactions and patients' reported interpersonal problems and outcome.

An idealized transference develops quite often, usually at the beginning of the treatment, whereas negative transference typically occurs later. Acute intense transference and countertransference feelings/reactions may serve as indicators of serious character pathology, such as Cluster B personality disorders. Devaluation presents one of the therapist's biggest challenges: conducting therapy and managing resistance with patients who force the therapist into very aggressive and uncomfortable experiences. When these situations arise, the therapist has a twofold task. He or she must tolerate the transference enough not to engage in a countertransferential enactment. Additionally, from this vulnerable vantage point, he or she must help the patient understand both the meaning of and the consequences of devaluations of the therapist[19].

Table 1. Examples of some kinds of transference and plausible reactions of the therapist.

Type of transference	Examples of typical thoughts	Emotional reactions	Behavior	Useful therapeutic reaction
Moderate positive	The therapist wants to help me, understand me, is human and helpful, doing his job very well – so he help me	Nice tune	Cooperation, willingness to do homework	None
Admiring – Independent	Therapist is great like me. It's nice to meet with such a person, he can advise me with something, of course I can only help myself.	Nice tune or euphoria at meetings	Enjoy discussing with a therapist opposite positions, slightly competes, homeworks alternately preparing and not doing, emphasizes freedom of relationship	Open narcissistic attitude, discuss the "advantages and disadvantages" in life, "normalization "- neither of us is more unique than any other man, insist on homework
Admiring-dependent	Therapist is great. It is only himself, who is able to help me - I am lost without him. I can do anything with his support.	Euphoric tune changes with anxiety about being left	Try to put themselves into therapist's place, They bring presents, flatter, do their best to do the homework- need affirmation that they did it right, often require advice, explanation, support.	To strength independent behaviors, assign exercises such as "independent steps" for homework, discuss the dependent thoughts and go through the advantages and disadvantages of the dependent attitudes. Advice, affirmation and further explanation should not be given. Exposition to independent activities and tasks.
Erotic	Therapist is a perfect (ideal) partner. Relationship with him will save me, it would be wonderful to be with him, make love with him, to flow into one another (be in harmony with him)	Crush on the therapist, maybe the depersonalization or the "trance" during the contact with the therapist	Flirting or shy withdrawal, wearing showy clothes, amorous looks towards the therapist, filtration of the information, i.e they need to look better than they really are.	It is not necessary to solve it, as far s it does not affect the therapy. It flows away spontaneously. In case it blocks the therapy- open the theme as a problem, go through the thoughts and realize their influence on the behavior, discuss sources of this need" in the past, express clearly and in a sensitive way his/her own attitudes in the therapeutic relationship.
Shy (apprehensive) -distrustful	Therapist can comment upon me, can hurt me, he is strong when he finds out who I am, he will refuse me and will make fun of me and will reject me	Fear, anxiety, shame	Problems with an eye contact, censorship of what he /she is willing to share with the therapist. Does homework anxiously, often speaks about being worthless, looks for apologies and explanations, rationalization, unable to trust.	Often goes away during the therapy, as the trust is obtained during the therapy, if it persists as a problem, it should be opened as a problem, work with the thoughts, look for the possible causes of these attitudes in the past, the adaptive and maladaptive aspects, planning the expositions with the aim of showing the courage to behave in an open way
Aggressive	I must show my power (dominance) otherwise I will be deprived of my freedom and discernment. It is either him who wins or me. How dares he! I must put him down!	Anger, hate, fear	Aggressive voice as well as his look, verbal rage, accusations, blaming, threatening, sometimes even the physical aggression.	Give the feedback, let the patient know we understand his feelings, let him express anger and derogation by the method of negative questioning and assertive approval or by the means of feedback, then discuss the attitudes lying behind that, go to their background, how they are expressed in the behavior, the advantages and disadvantages. When finally patient's tension decreases, the therapist should express his attitude by feedback (to take patient's behavior separately from his personality as such).

Table 1. (Continued)

Type of transference	Examples of typical thoughts	Emotional reactions	Behavior	Useful therapeutic reaction
Suspicious	Therapist does me wrong on purpose, abuses me for his own needs or for the needs of somebody else, he is against me, has hidden motives, does not play fair.	Anger, fear, feelings of threat	He withdraws, does not speak about himself or only superficially, can be aggressive indirectly, does not do homework, drops out or stops attending the therapy.	Give the feedback, discuss the situation openly, help to examine, where the sensibility comes from and to go through the relationships, where it also occurs. Mapping the sensible attitudes, their advantages and disadvantages, effects on the behavior. Experimenting with the confidence.
Competitive	Can't let him overtop me, I am better in many things than he is. I will show him, don't let him humiliate me.	Tension, changes of feelings of euphoria anger, envy, frustration, according to the subjective "score"	Secretly or openly competes, fiery discuss everywhere, where expects "competition", rationalizes non compliance in homeworks	Give the feedback for the specific situations, investigate the competitive thoughts and deeper attitudes, their sources situations, where they occur, to which behavior they lead to, advantages and disadvantages including their effects on the therapy..
Contemptuous	He cannot make it! He is weak, stupid, he is a fool, etc. How could he help me? I am the dominant one in our relationship	Contempt, impatience, anger	He despises the therapist, he cheapens what the therapist does, refuses to do the homework, drops out the lessons or stops attending the therapy	Give the feedback about the particular behavior, investigate thoughts and attitudes, find their origin, find how they work in different situations, the behavior they lead to, advantages and disadvantages for the life and relationships, what they mean for the therapy.
Jealous	He prefers the others, he does not care for me.	Ager, grief	Withdrawal or regrets, sometimes outbursts of anger, measuring the time of sessions (others and himself), monitoring of manifestation of favor (himself and others).	Ask for the thoughts related to the sorrow (harming the therapeutic relationship), then help with opening the thoughts, emotions and behavior concerning the anger. Go through the reasons in the past (place in the family as a sibling) and how they affect the behavior, emotions and relationships in various life situations, advantages/ disadvantages for the relationships and life and for the therapy.
Possessive	He is here for me, he has to be there for my disposal anytime.	Feelings of euphoria change with anger according to the behavior of the therapist.	He domineers, calls very often, does not visit the therapist at the time they agreed on and is angry when the therapist is not on disposal. He blames or is verbally aggressive.	Investigate thoughts and attitudes, find their origin in the past (the need of possession instead of the fear of being left by someone), thoughts, emotions and behavior in various relationships including the therapeutic one, advantages and disadvantages.

COGNITIVE ASPECTS OF TRANSFERENCE

Techniques of explicit formulation are included in cognitive behavioral therapy. Having a clinical formulation that is shared with a patient can help maintain the therapeutic alliance during difficult reenactments[1,20]. In CBT, especially in schema focus therapy, therapists use operationalized core schemas and beliefs as the focus of therapy, targeting transference and maladaptive interpersonal patterns. Developing a CBT case conceptualization of patients is recommended for treating each patient in CBT[4]; cognitive behavior therapists examine the thoughts, feelings, and behaviors related to a wide range of situations (including reactions to the therapist) and relevant childhood experiences to understand the underlying core beliefs and conditional assumptions of each patient[12].

Transference interpretation has remained a core ingredient in psychodynamic tradition, despite limited empirical evidence for its effectiveness. In the field of psychoanalysis, the technical use of transference interpretations versus other interpretations has been intensively debated over a period of 100 years. Despite this fact, the research base remains very limited and inconclusive. Only one of eight naturalistic studies has reported a positive correlation between transference interpretations and outcome[21]. There is no study on the efficacy of using transference discussions in CBT. The goal of transference interpretation is sustained improvement in the patient's relationships outside therapy. It seems to be especially important for patients with long-standing, more severe interpersonal problems. Although the central tool of CBT is not interpretation of transference, automatic thoughts and feelings related to interactions with the therapist are very much within the scope of exploration and may provide valuable opportunities for testing and modifying dysfunctional automatic thoughts[2,7]. One of the more common mistakes in CBT, is moving too quickly away from the emotions being expressed about the therapist or the therapy, and failing to sufficiently attend to this rich opportunity for further understanding the patient[16].

Tact and timing in the exploration of transference reactions are paramount. At the "macro" level of case analysis, formulation represents conceptualization at the level of whole treatment. Case formulation was initially developed in relation to psychodynamic approaches[22] and shown to be a replicable procedure. Recent work has included explicit formulation techniques in cognitive therapy[4].

Transference is also influenced by the actual behavior

of the therapist. Explicit discussion of the patient's ongoing relationship with the therapist is compelling when it is accurate. Focus on the transference makes it possible for the patient (and therapist) to become directly aware of the distinction between reality and fantasy in the therapeutic encounter. However, in brief therapy, transference interpretations may be too anxiety-provoking.

COUNTER-TRANSFERENCE

Counter-transference occurs when the therapist reacts in complementary fashion to the patient's transference. Attention to emotional reactions of both patient and therapist is a fundamental component of cognitive behavioral therapy, especially during the process of therapy with difficult patients. Despite the manualization of treatment and emphasis on techniques and pharmacotherapy, countertransference exists. No therapist is free of countertransference. To guide patients effectively in discovering their thoughts and expressing their feelings, the therapist needs to have a foundation of skills for recognizing, labeling, understanding, and expressing his/her own emotions[16]. To understand our own limitations, our own resistance to change, is necessary to discover more about the patient and ourselves; as we learn how the patient's behavior affects our own countertransference, we are also learn about how the patient affects others[23]. Rather than having no feelings, or being an expert at repression, the cognitive therapist is attuned to personal emotions that might affect the therapy environment. Just as the therapist would encourage a client to do, cognitive behavioral therapists use awareness to their own physical sensations and subtle mood shifts as cues, suggesting the presence of automatic thoughts. Any changes in the therapist's typical behavior might signal an emotional reaction and associated automatic thoughts, such as talking in a commanding (or hesitating) tone of voice, increased frequency of thoughts about a client outside sessions, or perhaps avoidance of returning a client's phone call or tardiness in starting or ending a session.

TYPES OF COUNTER-TRANSFERENCE

Betan et al.[24] studied a national random sample of 181 psychiatrists and clinical psychologists in North America. Each completed a battery of instruments on a randomly selected patient in their care, including measures of axis II symptoms and the Countertransference Questionnaire, an instrument designed to assess clinicians' cognitive, affective, and behavioral responses in interacting with a particular patient. Factor analysis of the Countertransference Questionnaire yielded eight clinically and conceptually coherent factors that were independent of clinicians' theoretical orientation: 1) overwhelmed/disorganized, 2) helpless/inadequate, 3) positive, 4) special/overinvolved, 5) sexualized, 6) disengaged, 7) parental/protective, and 8) criticized/mistreated. The eight factors were associated in predictable ways with axis II pathology. An aggregated portrait of countertransference responses with narcissistic personality disorder patients provided a clinically rich, empirically based description that strongly resembled theoretical and clinical accounts. Countertransference patterns were systematically related to patients' personality pathology across therapeutic approaches, suggesting that clinicians, regardless of therapeutic orientation, can make diagnostic and therapeutic use of their own responses to the patient[24].

In some cases, the focus on the patient's problems may allow the therapist to compartmentalize and avoid his/her own personal problems or allow the therapist to displace his/her conflicts with others onto the patient[23]. Some people are attracted to being therapists because it allows them a sense of competence, superiority, and apparent efficacy. This illusion of competence may allow the therapist to unconsiously pursue other goals, such as the need to have power or control, or the need to compartmentalize, intellectualize, and isolate oneself from one's own problems.

Table 2. Examples of the counter-transference and possible strategies for a change.

Type of counter-transference	Examples of typical thoughts	Emotional reactions	Behavior	Strategies of change
Moderate positive	I like him, he is nice, good cooperation with him, he will do well.	Nice tune	Cooperation, support, empathy	None
Admiring	That person is special (especially beautiful, original, intelligent, etc.)	Admiration, fascination	Therapist does not make appropriate examinations, does not conduct the therapy. Possible non-compliance of the patient is deprecated, does not require patient's homework, tends to talk about the exceptional properties of the patient	Clarify own attitudes, their background, the effect on the behavior, advantages + disadvantages for the therapy. Supervision needed. "Normalization of the therapy": conduct the same way like the others. In case that the behavior is impossible to change and make a standard therapy, necessary to open that problem with the patient or the patient should change the therapist
Over-protective	He cannot make decisions on his own, Leeds help, advice, it will be my fault, if something wrong happens to him.	Fear, insecurity	He gives advice, protects, ensures, takes control over the patient, does not allow patient's independent decision making, doubts patient's abilities	Clarify own attitudes, their background, the effect on the behavior, advantages + disadvantages for the therapy. Supervision needed. Stop the directive leading of the therapy, let the patient plan things, stop ensuring. Otherwise the patient should change the therapist.

Table 2. (Continued)

Erotic	He/ she is attractive. I would feel nice with him/her. The only problem is his/her lack of tenderness (sex, attention). Has sexual dreams (imagination) about the patient.	Fascination, "trance" or depersonalization during the time they meet	He flirts, is over-protective, "unwilling" touches, speaks often about sex, offers "sexual therapy" in the worst case and has an affair with the patient	Stop rationalization of the seductive behavior, stop it completely, admit own counter- transference, find supervision. Realize own motives, their background, the effect on the behavior, advantages + disadvantages for the therapy Otherwise the patient should change the therapist. Even after the change, the therapist should not have a sexual affair with the patient.
Apprehensive	He can hurt me, make fun of me, rouse me, show me I am worthless, stupid, etc.	Fear, anxiety, shame.	He speaks quietly, cannot keep the distance, the leading of the therapy leaves on the patient, is not active in the therapy (he calls it "empathic leadership". He is afraid to say what he thinks, does not discuss the alternatives with the patient	Work on the self-confidence and self-acceptance, help the patient to process the transference reaction. Supervision always needed. Otherwise the patient should change the therapist
Aggressive (invasive)	He is a psychopath, an ignorant person (does not try hard enough, wants only advantages, secondary benefits, etc). He is annoying. I will show him!	Anger, resonance	He moralizes, preaches, minimizes the needs of the patient, does not have time for the patient. He is rude to the patient, yells at him.	Realize his own aggressive attitudes and behavior, stop to deny or rationalize them. Clarify their background, the effect on the behavior, advantages + disadvantages for the therapy. Otherwise the patient should change the therapist.
Distrustful	What does he want actually? He has some hidden intentions against me!	Apprehension, tension, anger	Withdrawal, only "formal cooperation with the patient, waiting for hidden motives, tries to cancel the therapy	Work on the self-confidence and self-acceptance Supervision needed. To process own attitudes, their origin and effects. If necessary, let the patient chase the therapist
Competitive	Do not let him think, he will overtop me.	Tension changes with the pride (vanity)	Competition with the patient in the opinions, in "who is right", prides himself, he is not very supportive, empathic	Work on the self-confidence and self-acceptance Supervision needed. To process own attitudes, their origin and effects. If necessary, pass the patient to other therapist
Derogatory	He his a jackass, weakling, dumb, hysterical person, etc.) I am fed up with him bored,, I wish he would not annoy me.	Contempt, boredom, anger, vanity	He gives conceptive advice minimizes the attitudes and problems of the patient, make fun of him, does not have the time for him, is very inpatient, does not let the patient finish what he wanted to say, does not listen properly.	To work on his/her own relationships, attend the psychotherapeutic training or to attend a new one in case experience from the previous is not sufficient enough. To process own attitudes, their origin and effects. If necessary, pass the patient to other therapist

Therapist countertransference feelings may be informative about the entire treatment process of the patient. It has generally been recognized that countertransference vicissitudes play an extremely important role in psychotherapy of patients especially with hypochondriacal and borderline patients, at one or another phase of the treatment[25]. At critical points in the development of transference and countertransference and their interaction, the therapist's recognition of and capacity to deal with countertransference issues become crucial to the treatment progression. According Leahy[23] typical problems in countertransference include following:

- Ambivalence about using techniques because of fears of alienating the patient;
- Guilt or fear over the patient's anger;
- Feeling of inferiority when working with narcissistic patients;
- Discomfort if the patient is sexually attractive;
- Inability to set limits on sexually provocative or hostile patients;
- Overextending therapy sessions;
- Lack of assertion in collecting fees or enforcing policies;
- Inhibition in taking an adequate sexual history;
- Anger at patients who make phone calls between sessions;
- Catastrophizing the issue of hospitalizing a patient.

When patients present with issues such as abandonment, dependency, devaluation, demandingness, sexual preoccupations, abuse, betrayal, or exploitation of others, they may arouse your own feelings and vulnerabilities about these issues[23].

COGNITIVE ASPECTS OF COUNTERTRANSFERENCE

The therapist uses all of the automatic thought distorsions in countertransference: "This patient is hysterical." (labeling); "He wants only secondary gains." (mind reading); "He'll never get better." (fortune telling); "He's made no improvement." (all-or-nothing thinking); "It's my mistake the patient is no better." (all-or-nothing thinking); "He should do homework" (shoulds); "He won't get better." (overgeneralization)

Countertransference schemas

The therapist's individual schemas will affect the countertransference.There are probably as many schemas as there are adjectives to describe people. Many therapists display more than one schema. Moreover some patients are more likely to activate certain schemas than others. Leahy[23] describes these therapist schemas:

- *Demanding standards:* Anankastistic or perfectionistic therapists often view patients as irresponsible, self-indulgent, and lazy. They believe that the expression of emotions, or even uncertainty, is devastating. They have difficulty expressing warmth and empathy toward patients and place a great deal of emphasis on "logic" and "rationality". The patient may feel that the therapy is simply an opportunity for the therapist to show that he or she is smarter than the patient. Perfectionistic therapists may attempt to compensate his/her underlying feelings of incompetence and worthlessness by demanding perfect performance from self and patient. A typical sequence of automatic thought can be: "My patient is not getting better → I'm not doing my job → I'll be exposed as a fraud → I'm a failure → I can't accept any failure in myself." In some cases the therapist with demanding standards can compensate for his/her perfectionism by demanding more and more from the patient.
- *Abandonment*: the therapist with an abandonment schema will be worried that if he/she confronts the patient, then the patient will leave the therapy. Premature termination of the therapy is interpreted as a personal rejection of the therapist. Therapists concerned about abandonment issues can behave in many different forms that reflect countertransference: for example on the one hand, excessive caretaking of the patient, or on the other, avoidance of entering into a meaningful therapeutic relationship. Excessive caretaking takes the form of trying to protect the patient from any difficulties and taking on the patient's problems as the therapist's own to solve. Therapists who avoid attachment on the other hand, often focus more on superficial techniques than on more meaningful personal issues. This type of therapist avoids difficult topics with patients and refrains from using anxiety-provoking interventions, such as exposure techniques. They often personalize the patient's lateness, failure to show up for a session, or lack of interest in therapy. Patient's resistance can be seen as a personal rejection.
- *Special, superior person*: the narcissistic therapist views therapy as an opportunity to show of his/her special talents. Therapy with the resistant patient may begin with grandiose hopes, expressed by the therapist that the patient has finally found the "right therapist". The therapist's investment in his/her own image as being a special, superior therapist may result in his/her siding with the patient to vilify all the other therapists who have "failed" the patient. This therapist feels entitled to having the cooperation and adulation of the patient. This may result in the therapist encouraging boundary violations by the patient or, in some case, the therapist himself/herself may initiate these boundary violations. As the therapeutic relationship unfolds – if the patient does not make rapid progress – the narcissistic therapist may grow bored with, angry at, or punitive towards the patient. He may label the patient "He's a borderline, histrionic, paranoid, hypochondriacal..." Rather than empathize with the patient's understandable frustration with lack of progress, the therapist may turn on the patient, blaming the patient for a lack of desire to improve. To modify the narcissistic perspective, one needs to ask one question: "What would your life be like if you had to walk in the shoes of this patient?"

- *Need for approval*: The "pleasing" therapist may be highly skilled in showing empathy for the patient. He/she wishes to make the patient feel good regardless of what is going on, is averse to any expression of anger or disappointment by the patient. The warmth and empathy of such a therapist are much appreciated by many patients but he/she has difficulty recognizing that borderline patients are very angry. This type of therapist will avoid raising questions about the patient's substance abuse, anger, resistance, and self-defeat. These topics are viewed as too disturbing for the patient, and therefore as not appropriate. Patients may act out by missing sessions, showing up late, or not doing homework, but the high-need-for-approval therapist, who does not want to cause a "conflict" communicates the idea that acting-out behavior is acceptable. One therapist found it difficult to make the decision to hospitalize a suicidal patient because of his/her concern that the patient would get angry with him/her. The therapist may find that the patient's anger is difficult to tolerate. He/she can personalize the patient's behavior and viewed the patient's disapproval as a sign of his/her own failing. His/her assumption was, "If the patient is angry at me, it means that I failed."

SELF-CORRECTION OF COUNTER-TRANSFERENCE

By anticipating and paying attention to such countertransferential responses, CBT affords the therapist the opportunity to recognize and manage such responses, which reduces the therapist's risk of retaliatory acting out[7].

Throughout the process of providing therapy, in addition to tending to the patient's expressions, the therapist has to make an effort to monitor his/her reactions to the content of the sessions. Therapist must take special care to recognize his/her strong emotional reactions to patient, both positive and negative; this is an opportunity to ask him/herself how much of what the patient is processing matches the therapist s prior experiences or preexisting opinions.

The therapist monitoring his/her (positive and negative) feelings, must be aware of these reactions:
- Dreading or happily anticipating session with the patient;
- Having exceptionally strong hateful or loving feelings towards a patient;
- Wanting to end sessions early or extend sessions;
- Strongly wishing for or dreading termination.

The first step in managing counter-transference is the therapist recognizing that his/her feelings toward a patient are unusually strong, either positive or negative. It is use-

ful to take some time, perhaps outside the therapeutic environment, to patiently ask some introspective questions:

- What are my emotional reactions to this patient?
- Are they somewhat exaggerated?
- What is making me like or dislike this patient?
- What issues do I want or not want to discuss with this patient?
- What is making me feel uncomfortable?
- What were some signs of the patient's pathology that I had missed? What was it about me that made me miss them?

A second step may involve seeking out consultation with a supervisor to help delve deeper into addressing and potentially resolving the source of strong counter-transference feelings.

E.g. the therapist may find himself/herself frustrated, angry, anxious, or threatened by the patient's demand for validation. With the work with own thoughts he/she can recognize for instance[23]: "This patient doesn't really want to get better. All she wants to do is whine. She's keeping me from getting my job done. I'm going to look like I'm incompetent because she won't do what she should do. This patient is just irrational. She shouldn't be irrational." It is immensely stress reducing and helpful to the patient's therapy when the therapist can identify and challenge these negative countertransference thoughts. Challenges to these thoughts include the following: "It's irrational to think people should be rational all the time." "All of us need validation some of the time." Reflecting, caring, showing curiosity and respect, and being a good listener are interventions."

In order to examine the countertransference, the therapist should examine the kinds of life problems that he/she typically has. Is he/she someone who is concerned about rejection or abandonment? Then he/she should examine how these issues arise in his/her contact with patients. Is he/she someone who always has to be "right"? Then he must examine how he/she may be trying to defeat patients in debates, and thereby invalidate them. Is he/she someone who is afraid of failing, because he/she thinks that success or failure indicates how worthwile he/she is? Then he/she must examine how he/she may be afraid of dealing with difficult patients or afraid of taking chances in therapy.

The way the therapist views or deals with therapy-related thoughts and emotions may need some cognitive restructuring to reduce intensity of negative affect or to maintain adequate focus on therapy goals and objectives[16]. It may be useful to contront any fears about therapist emotions being "mistakes" or indications of failure in therapy and instead focus on ways of understanding the emotional antecedents. Therapist reactions can stem from a number of sources, including cultural of value-related beliefs, the therapist's view of his/her professional role, and unique learning history, as well as from the interactions with the patient's problematic behaviors[26]. The therapist can also use a self-directed inquiry of thoughts about a session, a situation or working with a particular patient or problem and log these thoughts into a dysfunctional thought record.

Especially in preparing to work professionally with patients suffering from personality disorders, hypochondria or somatoform disorders, the therapist needs to be especially careful to be nonjudgmental. Once the therapist has made the diagnosis, it is much better to avoid labels and think in terms of beliefs, core and conditional schemas, predictable reactions, behaviors and so forth. By trying to put him-/herself in the patient's shoes, perhaps imaging him-/herself with the same set of sensitivities, sense of helplessness, and vulnerability – the therapist can better understand the patient. At the same time, the therapist has to be on guard not to become so involved with the patient's problems that objectivity is lost.

COUNTER-TRANSFERENCE AND SUPERVISION

Having a formulation shared with the patient can help maintain the therapeutic alliance during difficult re-enactments; or, in supervision, help understand potential re-enactments[27]. Ongoing discussion of the therapy with colleagues and with supervisor is valuable (even for experienced therapists) and is built into those therapies that have been empirically validated[28]. Such discussions enhance the therapist' s ability to clarify the patient's transference and contain counter-transference anger and resentment[29].

Table 3. Therapist's Dysfunctional Thoughtr Record[16].

Situation	Emotion	Automatic thoughts	Rational response
Patient arrives late; persists with dramatic story-telling; breake into sobs when I redirecte to agenda setting.	Frustrated Disappointed Uncertain Embarrassed	This patient will never get is! We are making no progress using cognitive behavioral therapy! I don't know what to do next. I must be ineffective with the approach.	Contempt on my part will not help, so I could avoid such eternalized judgments and be more sympathetic. She is showing more skill in labeling affect, and identifying thoughts. Also, I'm focusing on the importance of making a list when her obvious priority is interpersonal support. I need to respect her values, help her learn to define problems, and not give up. Just because I feel uncertain does not mean I am ineffective, or have commited any shameful action. My discomfort comes from believing all patients must change quickly, and if they don't it's my fault. Does it make sense that an effective therapist "never" feels uncertain? I can brainstorm some options to try next.

Supervision should support the therapist, give another perspective to problem-solve difficult clinical dilemmas, bolster theoretical understanding to comprehend the patient's current issues, and assist the therapist in maintaining a benevolent, caring, and curious attitude to the patient's vicissitudes. Effective supervision related to the therapy should provide a safe place for the therapist to disclose feelings and attitudes[28].

Appreciation for the concept of transference might inform the cognitive behavior therapist's understanding of a patient's dysfunctional automatic thoughts and feelings.

CONCLUSION

The literature shows and it is our experience that both transference and counter-transference issues should be examined carefully and openly in CBT and must be an integral component of the complete management of each patient undergoing CBT. Analysis of transference aims to improve interpersonal functioning. Transference elaborations in CBT seem to be especially important for patients with long-standing problematic interpersonal relationships. Specifically, those patients who need to improve the benefit the most.

Countertransference can be one of the most useful tools in helping patients by providing a window into the "real-world effects" that the patient has outside the therapy. This can be helpful in diagnosing his/her problem and helping the patient understand how his/her behavior may affect others.

REFERENCES

1. Goin MK. A current perspective on the psychotherapies. Psychiatric Services 2005; 56:255–257.
2. Beck JS. Cognitive Therapy: Basics and Beyond. New York, Guilford 1995.
3. Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive Therapy of Depression. New York, Guilford 1979.
4. Persons J. Cognitive Therapy in Practice: A Case Formulation. New York, WW Norton 1989.
5. Gluhoski V: Misconceptions of cognitive therapy. Psychotherapy 1994; 31:594–600
6. Cutler JL, Goldyne A, Devlin MJ, and Glick RA. Comparing cognitive behavioral therapy, interpersonal psychotherapy, and psychodynamic psychotherapy. Am J Psychiatry 2004; 161:1567–1573.
7. Young JE, Weishaar ME, and Klosko JS. Schema Therapy: A Practitioner's Guide. New York, Guilford 2003.
8. Frank JD, and Frank JB: Persuasion and Healing. A Comparative Study of Psychotherapy. Baltimore, Johns Hopkins University Press 1991.
9. Joyce AS, Piper WE. The immediate impact of transference in short-term individual psychotherapy. Am J Psychother 1993; 47:508–526.
10. Wampold BE. The Great Psychotherapy Debate: Models, Methods, and Findings. Mahwah, NJ, Lawrence Erlbaum Associates 2001.
11. Linehan MM. Dialectical behavioral therapy in groups: Treating borderline personality disorders and suicidal behavior. In: Brody CM (ed), Women in groups. New York, Springer 1987.
12. Sareen J, and Skakum K. Defining the core processes of psychotherapy. Am J Psychiatry 2005; 162:1549.
13. Giesen-Bloo J, van Dyck R, Spinhoven P, van Tilburg W, Dirksen C, van Asselt T, Kremers I, Nadort M, Arntz A. Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs transference-focused psychotherapy. Arch Gen Psychiatry 2006; 63:649–658.
14. Knapp H. Therapeutic Communication. Developing Professional Skills. Sage Publications, Los Angeles 2007.
15. Breuer J and Freud S. Studies on hysteria. London: Hogarth Press 1955 (Original work published in 1895).
16. Beck AT, Freeman A, Davis DD and Associates: Cognitive therapy of Personality Disorder. The Guilford Press, New York 2004.
17. Hoffart A, Hedley LM, Thornes K, Larsen SM, and Friis S. Therapists' emotional reactions to patients as a mediator in cognitive behavioural treatment of panic disorder with agoraphobia. Cogn Behav Ther 2006;35(3):174–82.
18. Rossberg JI, Karterud S, Pedersen G, and Friis S. Specific personality traits evoke different countertransference reactions: an empirical study. J Nerv Ment Dis 2008; 196:702–708.
19. Robbins B. Under attack: devaluation and the challenge of tolerating the transference. J Psychother Pract Res 2000; 9(3):136–141.
20. Spinhoven P, Giesen-Bloo J, van Dyck R, Kooiman CG, and Arntz A. The therapeutic alliance in schema-focused therapy and transference-focused psychotherapy for Borderline Personality Disorder. J Consul Clin Psych 2007; 75:104–115.
21. Høglend P. Analysis of transference in psychodynamic psychotherapy: a review of empirical research. Can J Psychoanal 2004; 12:279–300.
22. Luborsky L. Principles of Psychoanalytic Psychotherapy: A Manual for Supportive-Expressive (SE) Treatment. New York. Basic Books 1984.
23. Leahy RL. Overcoming Resistance in Cognitive Therapy. The Guilford Press, New York 2003.
24. Betan E, Heim AK, Zittel Conklin C, Westen D. Countertransference phenomena and personality pathology in clinical practice: an empirical investigation. Am J Psychiatry. 2005; 162(5):890–898.
25. Meissner WW. Notes on countertransference in borderline conditions. Int J Psychoanal Psychother 1982–1983; 9:89–124.
26. Kimmerling R, Zeiss A and Zeiss R. Therapist emotional responses to patients: Building a learning-based language. Cogn Behav Pract 2000; 7:312–321.
27. Margison FR, Barkham M, Evans C, McGrath G, Clark JM, Audin K, and Connell J. Measurement and psychotherapy: Evidence-based practice and practice-based evidence. Br J Psych 2000; 177:123–130.
28. Gunderson JG, and Links PS: Borderline Personality Disorder. A Clinical Guide. American Psychiatric Publishing, Inc. Washington 2008.
29. Gabbard GO and Wilkinson SM. Management of Countertransference With Borderline Patients. American Psychiatric Press, Washington, DC, 1994.

The working alliance and the transference/ countertransference relationship: Their manifestation with racial/ethnic and sexual orientation minority clients and therapists

CHARLES J. GELSO AND JONATHAN J. MOHR

University of Maryland, College Park

Abstract

Theoretical propositions are offered about the interrelations of three key concepts in psychotherapy: specifically, the working alliance, on the one hand, and both transference and countertransference, on the other. The role these concepts play, including the interactions between them, is extended to the areas of race/ethnicity and sexual orientation. We examine working alliance, transference, and countertransference considerations for therapy dyads in which at least one of the participants is a member of a racial/ethnic or sexual orientation minority group. Our hope is that this theoretical examination will generate further research and theory development on working alliance, transference, and countertransference where cultural factors such as race/ethnicity and sexual orientation are implicated in the therapeutic relationship.

Key words: Countertransference, Culture, Race/Ethnicity, Sexual orientation, Transference, Working alliance

The concept of working alliance in psychotherapy has received much attention during the past decade, both theoretically and empirically. Originating from classical psychoanalysis (e.g., Greenson, 1967), this concept has been extrapolated to virtually all theoretical approaches to treatment. Bordin (1979) was probably the first to make such an extrapolation, and since his seminal article, the working alliance (also called helping alliance, therapeutic alliance, or simply alliance) has been arguably the most vigorously investigated construct in psychotherapy (Gelso & Hayes, 1998).

Like working alliance, the constructs of transference and countertransference are historically situated within psychoanalysis. Freud's earliest and most influential theoretical statements on transference occurred shortly after the turn of the 20th century (Freud, 1905/1953, 1912/1959b). Likewise, his brief but incisive conceptions of countertransference originated as part of his early statements about psychoanalysis (Freud, 1910/1959a). Only within about the past decade or so have there been efforts to generalize these two interrelated constructs to approaches other than psychoanalysis and to study them empirically (see review by Gelso & Hayes, 1998).

Transference, for example, has been studied in both psychoanalytic treatments (Gelso, Hill, Mohr, Rochlen, & Zack, 1999; Luborsky & Barber, 1994; Patton, Kivlighan, & Multon, 1997) and nonanalytic treatments (Barber, Foltz, & Weinryb, 1998; Gelso, Kivlighan, Wine, Jones, & Friedman, 1997). Programmatic research has also emanated from social psychology, which has explored transference phenomena in interpersonal relationships outside of psychotherapy (see review by Andersen & Berk, 1998).

In recent years, the first author has developed a series of theoretical propositions about the operation of working alliance, transference, and countertransference beyond the confines of psychoanalysis, including propositions regarding the interrelations of these three constructs (Gelso & Carter, 1985, 1994; Gelso & Hayes, 1998). The present article is an extension of that work. A major purpose here is to extend the first author's theoretical statements about the patterns of reciprocal influence between the working alliance and the transference/countertransference configuration.

During the past decade or so, another body of literature in psychotherapy has been growing at a rapid pace. Coming under the umbrella of diversity and cultural topics, investigators have begun to examine how sociocultural constructs such as race/ethnicity and sexual orientation influence the therapy process (Helms & Cook, 1999; Pope-Davis & Cole-

man, 1997; Thompson & Carter, 1997). In this literature, central questions revolve around how therapy process and outcomes are affected by various racial and sexual orientation pairings of therapist and client in which at least one participant in the dyad is a member of a racial/ethnic or sexual orientation minority group. For example, what is the effect on process and outcome of having a White therapist and an African American client? A gay male therapist and a gay male client? An African American therapist and a Latina client? Within this body of literature, there is also a nascent interest in how the psychological qualities of therapist and client moderate the effects of racial and sexual orientation pairings. Constructs such as racial identity (Helms & Cook, 1999; Thompson & Carter, 1997) and gay and lesbian identity (Fassinger & Miller, 1996; McCarn & Fassinger, 1996; Mohr & Fassinger, 2000) exemplify these psychological qualities. Despite the recent attention to a range of diversity topics in general and to race/ethnicity and sexual orientation within the context of psychotherapy, little theoretical or empirical work has evolved regarding the development of working alliance and transference/countertransference in therapy dyads in which one or both participants are members of racial or sexual minority groups.

In the present article, we seek to bring together recent theoretical work on the therapy relationship with emergent thinking about psychotherapy involving racial/ethnic or sexual orientation minorities. We begin by extending Gelso and Carter's (1994) and Gelso and Hayes' (1998) conceptualization of how working alliance and transference/countertransference interrelate across diverse approaches to psychotherapy. Second, we examine how these interrelationships are expressed in the context of psychotherapy in which at least one participant is a member of a racial/ethnic or sexual orientation minority group. In doing so, we also consider the role of racial/ethnic and sexual orientation identity factors as moderators. An overarching aim is to analyze how broad sociocultural factors related to societal oppression get expressed in one way or another in the working alliance, the transference/countertransference configuration, and the interconnection between the two.

The Working Alliance and the Transference/Countertransference Configuration

Let us begin by offering working definitions of working alliance, transference, and countertransference. To do so, we provide a background summary of each construct: its history, current status, and empirical relationship to psychotherapy outcomes. As is perhaps true with all high-level constructs in the behavioral sciences, defining the three major constructs is no easy feat. Embedded in what seems to many to be the obscure and perhaps arcane language of psychoanalysis, clear (to say nothing about operational) definitions appear to be even more difficult to arrive at than is typical.

The Working Alliance

Although the roots of the concept of working alliance may be found in Freud's earliest writings (Freud, 1905/1953, 1913/1958) as well as in that of several other early analysts, it was Ralph Greenson (1965, 1967) who formally labeled and defined the construct and explored its central role in psychoanalysis. Outside of psychoanalysis proper, Bordin (1979) probably had the greatest influence in generalizing the working alliance concept so as to be meaningful in all forms of counseling and psychotherapy.

Drawing on the work of Greenson and Bordin, the first author has defined the working alliance as *the joining together of the reasonable self or ego of the client and the therapist's analyzing or "therapizing" self or ego for the purpose of the work* (Gelso & Carter, 1994; Gelso & Hayes, 1998). This definition relies on the psychoanalytic concept of the split ego (Sterba, 1934). The ego can be thought of as divided into two sides: an experiencing side and an observing side. The experiencing side allows the client to feel and experience; the observing side permits reasonable observation and understanding of that experience as well as of the external world. It is the latter—the reasonable side or ego—that is most implicated in the working alliance. This side fosters client and therapist joining together, despite internal forces that may operate against doing so (e.g., negative transference or countertransference), for the purpose of the work, which itself aims to help the client solve problems, gain insight, change behavior, etc. When the working alliance is sound, as Bordin (1979) noted, client and therapist tend to have a solid working bond. They also tend to agree on the goals of treatment as well as the tasks needed to accomplish those goals.

Although there are several different definitions and measures of what we call the working alliance, two elements seem to cut across all of these: collaboration and attachment. That is, the fundamental idea of the therapy participants collaborating in the process and forming a healthy attachment to one another seems to be present in virtually all conceptions and measures. Regarding the effects of the working alliance, a large body of empirical evidence points to the clear conclusion that the alliance has a significant positive impact on the outcomes of brief and longer term psychotherapy (Martin, Garske, & Davis, 2000). Particularly in brief therapy, the alliance that is formed within the first few sessions is predictive of outcome. (for a discussion of the particular ways in which alliance affects outcome, see Gelso & Hayes, 1998, pp. 34–35).

The Transference/Countertransference Configuration

Whereas there appears to be general agreement that the working alliance is a key relational factor in effective psychotherapy, the concepts of transference and countertransference remain controversial. Transference, in particular, may be one of the most controversial constructs in psychotherapy. Depending on one's theoretical orientation, it may be seen as a vital aspect of successful therapy, an irrel-

evant construct, or an artifact of the therapist's stance toward the patient. In all versions of psychoanalytic and psychodynamic treatments, transference is seen as a key concept. To other major theory clusters, such as humanistic-experiential or cognitive-behavioral, its importance ranges from modest to trivial. Within these general systems, there is often acknowledgment that transference occurs and occasionally needs to be dealt with; but it is rarely seen as central, and it usually does not need to be focused on directly.

Based on empirical studies, the first author theorized that transference occurs and is important in all forms of psychotherapy (Gelso & Carter, 1985, 1994; Gelso & Hayes, 1998) and suggested ways in which it operates in different treatments. Drawing on both classical (e.g., Greenson, 1967), interpersonal (e.g., Fromm-Reichmann, 1950), and postmodern intersubjective (Orange, Atwood, & Stolorow, 1997; Stolorow, 1998) theories, transference is defined as "the client's experience of the therapist that is shaped by the client's own psychological structures and past, and involves displacement, onto the therapist, of feelings, attitudes, and behaviors belonging rightfully in earlier significant relationships" (Gelso & Hayes, 1998, p. 51). Although the client's experience of the therapist is shaped by the client's structures and past, the therapist does contribute something to this experience (Gelso & Hayes, 1998). In this sense, no two therapists have exactly the same transferences from the same client; to a degree, all transferences are cocreated by therapist and client. At the same time, some core issues within the client become attached to any therapist who permits the transference to occur. In other words, as has been empirically documented, each client carries with him or her a very limited number (often one) of core conflictual relationship themes (Barber et al., 1998; Luborsky & Crits-Christoph, 1990; Luborsky, Popp, Luborsky, & Mark, 1994), and from these themes emanate the transferences to essentially any therapist with whom he or she works.

Regarding the effects of transference across therapies, there is some emerging but still very tentative evidence that transference has a positive effect when it is combined with insight on the client's part (Gelso, Hill, & Kivlighan, 1991; Gelso, Hill et al., 1999; Gelso, Kivlighan et al., 1997; Graff & Luborsky, 1977), and this effect seems to occur across different theoretical approaches to treatment. In the absence of client insight, however, transference (especially when it is negative) appears to have an adverse effect on outcomes. (It should be noted that, in virtually any conception of transference, reactions may be seen as existing on a continuum from positive to negative. At the midpoint in the continuum, some would conceive of transference as neutral with respect to valence.) We return to the effects of transference later in the article.

Regarding countertransference, although the term is a psychoanalytic one, it seems to us that there is a consensus that, at least in a general sense, the therapist's emotional reactions and issues are a significant consideration in any approach to treatment (Gelso & Hayes, 1998). The valuing of such relational factors has become notable even in therapies that were once decidedly nonrelationship oriented, such as the behavioral therapies (Cahill, Carrigan, & Evans, 1998). Despite the emerging agreement, it is important to understand that countertransference itself has been variously defined over the years. Gelso and Hayes (1998) delineated three fundamental definitions. The first (the classical psychoanalytic one) understands countertransference as the analyst's transference to the analysand's transference. This definition reflects Freud's belief that countertransference represented the analyst's inappropriate displacement of material from previous relationships onto the patient, and that it should be eliminated due to its likely pernicious effects on the analysis (Freud, 1910/1959a). The analyst's own personal psychoanalysis was seen as perhaps the best road to minimizing countertransference reactions.

A second conception of countertransference is often referred to as the totalistic definition. Here countertransference is seen as virtually all of the therapist's emotional reactions to the patient. Countertransference is viewed by totalists as inevitable and potentially useful if the therapist pays close attention to his or her inner workings in response to the patient. Through studying one's internal reactions, the therapist can gain insight to how the client affects others and how others react to the client. To many psychoanalytic therapists, the therapist's self-inspection of his or reactions can also yield understandings about the patients' early conflicts with transference sources (e.g., parents).

Criticisms of the classical definition revolve around its being too narrow. Similarly, complaints with the totalistic definition usually pertain to its being too encompassing. In an attempt to reconcile these definitional differences, we espouse an intermediate conception. Countertransference is defined as "the therapist's transference to the client's material, both transference and nontransference communications presented by the client" (Gelso & Carter, 1994, p. 297). In maintaining the view that countertransference represents the therapist's transference, this conception is narrower than the totalistic one. Countertransference is seen as rooted in the therapist's own unresolved conflicts and issues. Yet it may be a response to any material offered by the client, transference or otherwise.

Although the intermediate conception of countertransference locates this phenomenon in the therapist's unresolved issues (even while triggered by something in the client), countertransference is still seen as potentially invaluable if the therapist pays close attention to his or her inner workings and seeks to grasp how these relate to the client, the client's issues, and the client's enactments during treatment. Along with being potentially useful, countertransference has been viewed as inevitable and universal (Gelso & Carter, 1985, 1994; Gelso & Hayes, 1998). No matter how well analyzed a therapist may be, there will always be unresolved issues or what may be called "soft spots." These are often touched on

in the intimate and emotionally demanding experience of psychotherapy and require understanding and management. It is when the countertransference is not understood (i.e., the acted out countertransference) that it becomes pernicious. The understood countertransference can be extremely helpful.

It should be added that reactions that are not countertransference (e.g., realistic dislike of an abusive and arrogant patient where the therapist's feelings are not rooted in earlier conflict) are also important and need to be understood. We believe, however, that it is useful to differentiate these from countertransference, such that countertransference is seen as a phenomenon that is specific rather than representing any and all therapist reactions to the client.

In recent years, there have been increased efforts to study countertransference empirically (Gelso & Hayes, in press). Still, very few studies have attempted to link countertransference to outcome. In putting together the pieces of process studies that indirectly linked countertransference to outcome, Singer and Luborsky (1977, p. 449) concluded that "uncontrolled countertransference has an adverse effect on therapy outcome. Not only does it have a markedly detrimental influence on the therapist's techniques and interventions, but it also interferes with the optimal understanding of the patient." Hayes, Riker, and Ingram (1997) uncovered a direct link of countertransference to outcome, finding the expected adverse effect. Similarly, a recent study (Gelso, Latts, Gomez, & Fassinger, in press) revealed that therapists' ability to manage their countertransference resulted in more favorable therapy outcomes in a university counseling center.

The Interrelation of Working Alliance and Transference/Countertransference

What are the connections between working alliance, on the one hand, and both transference and countertransference, on the other? Are there salient patient, therapist, and treatment factors that moderate these connections? We begin with a general theoretical proposition about working alliance and transference, and then elaborate the more specific ways in which the two are interrelated.

Proposition 1: Working alliance and transference have a reciprocal impact on one another, with each influencing and being influenced by the other. If there is a foundational construct in psychotherapy, we propose that it is the working alliance. It is hard to imagine a therapeutic relationship in the absence of at least a "good enough alliance," and, as we have noted, considerable empirical evidence has accumulated to support the importance of the working alliance. Regarding the reciprocal relation of alliance and transference, let us first examine how transference affects the alliance and its formation. We look at this causal direction first because, in our view, transference precedes the formation of alliance in psychotherapy. That is, transference occurs from the first moment of contact between therapist and patient and even in fact prior to such contact in the form of preformed transferences. These transferences occur in response to the therapist of the patient's fantasies as the patient considers beginning treatment and in response to the initial phone contact.

Positive Transference and the Working Alliance

Early in treatment, positive transferences are very common, often in the form of projections onto the therapist of the wished-for good mother or good father that the patient never had and at times the all-good mother or father (as in idealizing transferences) (Kohut, 1977, 1984; for empirical support for positive transference early in therapy, see also, Gelso et al., 1997; Patton et al., 1997). Early positive transferences were referred to by Freud (1912/1959b) as "unobjectionable positive transferences." Freud suggested that they may aid the therapist in forming a sound initial working alliance. To the extent that they aid the work and are not intensely positive or eroticized, our observations suggest that such transferences are often not even seen as transferences, especially by nonanalytic therapists, who by definition are not on the lookout for signs of transference.

Although positive transference can be an aid to the working alliance, transference may be seen as an untrustworthy ally or what Greenson (1967) referred to as a treacherous ally. In fact, we would consider working alliances that are built solely on a foundation of transference to be pseudo-alliances. The problem with working alliances that are based on transference alone is that transferences can dissipate; in fact, they do tend to dissipate in the course of therapy, whether the treatment is brief or long-term therapy and whether it is psychoanalytic or theoretically heterogeneous (see studies by Gelso et al., 1997; Graff & Luborsky, 1977; Patton et al., 1997). When transference dissipates, a working alliance based on transference also tends to erode. The following case vignette exemplifies an eroding working alliance that was solely based on transference.

The client was a 27-year-old doctoral student in English literature. She experienced debilitating anxiety around work and interpersonal issues and was described early in treatment as a "deeply frightened woman, frightened of her aggressive impulses, frightened of deep dependency issues, and frightened of what may be a very shaky cohesiveness." After an initial series of sessions, her therapist, an experienced psychoanalytic therapist, noted that due to her fears "she uses every defense imaginable to discount and minimize her therapist." He went on to note, however, that she did have some insight into this pattern, and that at least a part of her "clicked" with the therapist. He noted a beginning working alliance, but said that perhaps the clicking was coming out of "some transference basis that he would pay for later in the work." As work progressed it became apparent that the client imbued the therapist with magical powers, while at the same time being disappointed that he would not solve her problems. The positive transference gradually eroded as the client's issues around primitive dependency emerged, along with a belief that no one could ever provide her with the safe haven she des-

perately needed but never experienced. Because no gen-
uine working alliance had formed, apart from that based in
the positive transference, the work ended with the client
feeling unhelped after a year of treatment in what at that
time was her fourth attempt at therapy.

In this example, what seemed like a working alliance real-
ly was a transference (again, pseudo-alliance seems apt), and
when the "magical transference" eroded nothing positive was
left in the overall therapeutic relationship. On the other hand,
we would maintain that a strong working alliance, one based
on realistic rather than transference perceptions of the thera-
pist, can be spurred on by positive transference. The dissipa-
tion of transference, then, would not destroy the therapeutic
relationship because the alliance would already have been so-
lidified by a realistic connection between therapist and client.

Negative Transference and the Working Alliance

Just as positive transferences can aid the working alliance,
early negative transference can not only threaten the alliance,
but also bring the therapy to an unsuccessful ending. The
client's negative perceptions of and feelings toward the ther-
apist, which are the defining features of negative transfer-
ence, can infect the working alliance, to use a medical anal-
ogy, and the disease can be fatal. The deleterious effects of
early negative transferences are especially notable under cer-
tain conditions. First, when the therapy is time-limited and/
or brief, there is often too little time for such distortions to be
resolved. This became clear to the first author in his empiri-
cal study of time-limited therapy (eight-session limit) at a
university counseling center. Semistructured interviews with
a small number of clients whose therapy outcomes were es-
pecially positive revealed that a positive early relationship,
marked by the near absence of negative transference, was a
striking characteristic of these clients' reactions to their ther-
apists (Adelstein, Gelso, Haws, Reed, & Spiegel, 1983). More-
over, in a series of studies comparing time-limited versus un-
limited counseling, the ability to form a positive relationship
very early in the work appeared to be a key factor in suc-
cessful outcome in time-limited treatment (Gelso & Johnson,
1983). In contrast, a retrospective study of successful long-
term cases suggested that the early stages of treatment were
just as likely to feature negative transferences as positive
transferences (Gelso et al., 1999).

A second condition under which the adverse effects of neg-
ative transference are especially notable occurs when the
therapist is unable to recognize that the patient's reactions are
from another time and place. Although the concept of trans-
ference emanates from psychoanalytic theory and is most
prominently addressed in that theoretical approach, and al-
though it is overwhelmingly clear that effective therapy can
and does occur in the absence of direction attention to trans-
ference by the therapist (Lambert & Bergin, 1994), we main-
tain that there are times when transference must be dealt with
if therapy is to be effective. Perhaps the most important time
is when there is strongly negative transference in the initial

stage of treatment. If the patient is not aided in understand-
ing the source of these negative reactions and does not come
to see that such reactions are not fitting of the therapist, it is
hard to see how the work can proceed effectively. When there
are such negative transferences, we believe that the therapist,
at a minimum, needs to work at understanding and helping
the patient understand where he or she is coming from.
Something in the therapist may well have triggered the reac-
tion, and that needs to be empathically understood and
owned; but the core of the negative reaction must be under-
stood as belonging more fittingly in an earlier time and place
in the patient's life. In sum, our proposition is that early neg-
ative transference, especially when it is intense, needs to be
understood as such if the therapy is to move forward and suc-
ceed. If the client does not come to understand the transfer-
ence on his or her own, in the context of an empathic ambi-
ence, then it behooves the therapist to make interventions that
aim at fostering such understanding (e.g., interpretations). If
such negative transference is not understood early on (e.g.,
in brief therapy) and erodes the working alliance, there is ev-
idence to indicate that transference interpretations in the lat-
er part of treatment not only do not help, but also may lead
to the premature termination of treatment (Piper et al., 1999).

The need to examine initial and intense negative transfer-
ence is illustrated in the following case vignette.

Within minutes of the initial greeting between therapist
and patient, the patient, a woman in her mid-20s, began to
verbally attack the therapist, indicating that she had a sense
that the therapist (also a female) was cold and mean. The
patient saw this in the therapist's eyes. When asked what
the therapist's eyes reminded her of, the patient revealed
that they looked strikingly like the eyes of the patient's
mother. An exploration of the patient's experience of the
mother's coldness ensued Although negative transference
reactions naturally were not eliminated, the exploration of
the basis for the patient's initial reactions helped solidify
the beginning alliance and prevent intense negative trans-
ference from poisoning the therapeutic relationship.

It should be added that often the patient's transference cre-
ates affects in the therapist that serve to confirm the patient's
feelings and perceptions. As an example, the patient who
feels the therapist harbors underling negative feelings for him
or her that may very subtly behave in a way during the hour
(e.g., respond in a passively critical manner) to stir up nega-
tive affects in the therapist. In such an instance, the initial per-
ceptions are still transference-based, but the therapist's reac-
tions (reality-based and/or countertransference-based) hook
into the transference. We suggest that in such instances the
therapist generally should acknowledge his or her affective
reaction and then help the patient explore where the patient
was coming from, how he or she served to stir up that reac-
tion in the therapist, and how this relates to early and present
patterns in the patient's life.

We have already implied the third condition under which

goals and tasks of treatment. Given this lack, the therapist felt anxious during the hour and handled the anxiety by vacillating between excessive sympathy and cool detachment, which itself made alliance formation difficult. The anxiety was rooted in more general fears of being insufficient, as was her way of handling the anxiety.

In this example, one can clearly see the reciprocal influence of alliance and countertransference. The lack of safety due to a poor alliance formation and the effects of this lack are also evident.

In addition to fostering a sense of safety in the therapist, the good working alliance allows the overall relationship to survive countertransference problems. Despite the fact that the alliance fosters a sense of safety and diminishes countertransference acting out, there is an extent to which countertransference is universal. The intimate exploration of another person's conflicts, in the context of a relationship that seeks to provide emotional help, is bound to touch on the therapist's personal and unresolved conflicts. This cannot *not* happen. What we can hope for is that such conflicts are grasped by the therapist and are not acted our during the hour. And yet here too there is bound to be a certain amount of seepage, wherein the therapist's issues emerge into the treatment. When the alliance is solid, however, not only is this seepage kept to a minimum, but also its potentially deleterious effects are minimized. In other words, because of the working alliance, the overall relationship can survive some degree of countertransference acting out, provided that such acting out is not too intense or extensive. Consider the following vignette from the first author's casework (see Gelso & Hayes, 1998, p. 144).

> The patient (a 35-year-old woman) and the therapist had worked together twice a week for 2 years in analytically oriented therapy. It was clear that they had developed a sound working alliance. However, during one phase of the work, each time the client discussed her difficulties with her children, the therapist seemed to respond unempathically, often offering unneeded suggestions and at times appearing even critical. The patient perceived the therapist's expressions realistically and responded with realistic hurt and frustration. She expressed these feelings to the therapist, along with her impressions of his reaction during one session, and the therapist became aware that, in fact, his conflicts surrounding his own parenting, as well as earlier issues with the parenting he received, were impeding the empathic process with this patient. Once the countertransference issues were grasped, the therapist was able to regain his empathic stance during the client's exploration of her parenting.

Here, the sound working alliance allowed the relationship to withstand the therapist's acting out of countertransference. Indeed, it permitted the patient openly to express her concerns about the therapist's behavior and to cue the therapist

that something was amiss. Of course, this positive effect of alliance on countertransference assumes that the therapist gains an understanding of the countertransference issues. To do so, the therapist must appreciate the importance of self-reflection, countertransference, and what we believe to be the vital importance of dealing with one's countertransference issues.

The Management of Countertransference

The preceding example leads us to our final suggestion about the connection of working alliance to countertransference. That is, if countertransference is not understood by the therapist or effectively managed in one way or another, it is likely to impede the formation of working alliance and erode an existing alliance. Ununderstood or unmanaged countertransference can produce irreparable ruptures in the alliance. In so affecting the alliance, unmanaged countertransference indirectly produces negative therapy outcomes. (In the path analytic sense, we have a direct effect of countertransference on outcome and an indirect effect in that countertransference affects outcome through its effects on alliance.) Although research bearing on these assertions is almost nonexistent, one recent study (Ligiero & Gelso, 2001) did discover that greater amounts of both positive and negative countertransference were related to poorer working alliances in brief therapy.

The first author and his collaborators have been involved in theoretical and empirical work on the topic of countertransference management. Theoretically, it has been proposed (Gelso & Hayes, 1998; Van Wagoner, Gelso, Hayes, & Diemer, 1991) that countertransference management relies on, and in certain ways is composed of, five factors or constituents: self-insight, empathy, self-integration, anxiety management, and conceptual ability. Briefly, self-insight refers to therapists' ability and willingness to look inward and understand themselves and their issues as they relate to the therapeutic relationship; empathy reflects the capacity to identify partially with the client (i.e., to climb into the patient's shoes and experience the world, to an extent, as the patient does, both in a cognitive and an affective sense); self-integration indicates the capacity to be separate from the patient, to understand where the therapist stops and the patient starts, without being distant from the patient; anxiety management relates to the therapist's ability to deal effectively with anxiety in the treatment hour (it does not imply an absence of anxiety); and conceptualizing ability refers to the conceptual understanding of the dynamics of the relationship and how the patient's and therapist's feelings and issues inform those dynamics. It should be noted that these five factors have been conceptualized to have trait-like qualities in that they are a part of the therapist and tend to be exhibited across patients. At the same time, it also true that these factors will vary to an extent from client to client and in this sense may be construed as states as well as traits.

In a series of empirical studies (done in both the laboratory and the field), each of these five factors, separately and

combined, has been found to be related to treatment process (Friedman & Gelso, 2000; Gelso, Fassinger, Gomez, & Latts, 1995; Hayes & Gelso, 1991, 1993; Hayes, McCracken, Mc-Clanahan, Hill, Harp, & Carozzoni, 1998; Latts & Gelso, 1995; Peabody & Gelso, 1982; Robbins & Jolkovski, 1987; Sharkin & Gelso, 1993) and to treatment outcome (Gelso et al., in press; Hayes et al., 1997; Rhodes, Hill, Thompson, & Elliott, 1994). Interestingly, these five factors, in combination, have been found to relate negatively to the expression of countertransference behavior in a given hour or what we have been referring to as the acting out of countertransference (Friedman & Gelso, 2000).

We now shift our attention to how working alliance and transference/countertransference are experienced and expressed in treatment dyads in which one or both participants are members of a racial/ethnic or sexual orientation minority group. Here we are mapping new territory. There is a paucity of theoretical work on this topic, and we are aware of few empirical investigations directly pertinent to our focus.

The Working Alliance and the Transference/ Countertransference Configuration in Cultural Pairings

In developing our thinking about how working alliance and transference/countertransference are moderated by racial/ethnic and sexual orientation factors, we were sharply aware of our own status as nonminority males. Despite the fact that we have worked with racial/ethnic minority patients and have conducted research on diversity related topics (e.g., Miville et al., 1999; Fuertes & Gelso, 2000; Mohr & Rochlen, 1999), we felt that it would be profitable to explore the topic with recognized experts in the area of race/ethnicity and the psychotherapy relationship. With this in mind, we conducted semistructured telephone interviews of approximately 1 hour in duration with Donelda Cook, Cheryl Thompson, and Peggy Rios. Each of these psychologists is known to be an expert in the area of race and ethnicity, and each has been involved in research, theory, and training related to issues of race and ethnicity in psychotherapy (see, e.g., Helms & Cook, 1999; Rios, Corbett, Pannu, Kanitz, & Miville, 1998; Thompson, 1995). We believed that interviewing a recognized expert on the topic of sexual orientation and the therapy relationship would also be helpful, although we have also had experience in working with gay, lesbian, and bisexual patients and one of us identifies as a gay man. We thus conducted a 1-hour semistructured phone interview with Jack Drescher (see Drescher, 1998).

Several questions were mailed to the experts in advance of the phone session, with the suggestion that they reflect on them prior to the interview. These questions were as follows.

1. In what ways does race, ethnicity, and sexual orientation play a role in the development of transference, development of working alliance, and interplay between transference and alliance?

2. How do the dynamics related to transference and alliance differ in your work with clients of different races, ethnicities, and sexual orientations?
3. How might these dynamics differ for therapists who are different from you in race, ethnicity, or sexual orientations?
4. Describe one or more cases that help to elucidate the ways in which race, ethnicity, and sexual orientation affect the interplay between transference and alliance or the development of transference and alliance.

As we began our theoretical exploration, we were certain that the relationships of alliance to transference/countertransference, as postulated earlier in this article, would be influenced and altered by race, ethnicity, and sexual orientation. And yet in no way were we or the experts able to articulate such a moderating effect. Thus, we offer a proposition that we fully expect to be controversial. We offer this proposition with the knowledge that it requires further theoretical and empirical exploration, and we welcome such exploration.

Proposition 3: The interrelations that we have proposed between the working alliance and the transference configuration are fully generalizable to all patient–therapist pairings, including those in which at least one participant is a member of a racial/ethnic or sexual orientation minority group. Whereas we do not believe that racial/ethnic or sexual orientation minority (RSM) status moderates the relations of working alliance and transference/countertransference, we do propose that RSM status does itself influence the development of both working alliance and transference/countertransference. Indeed, writers have noted the importance of attending to RSM status in understanding the dynamics related to these components of the therapy relationship (e.g., Chin, Liem, Ham, & Hong, 1993; La Roche, 1999; Morrow, 2000). Thus:

Proposition 4: The particular pairings representing RSM status influence the specific ways in which working alliance, transference, and countertransference emerge and unfold in all forms of psychotherapy. Before exploring the ways in which RSM status enters these aspects of the therapy relationship, some clarification of terms is necessary. The concepts of transference and countertransference take on added meanings when race, ethnicity, and sexual orientation enter into the relationship. We examine these added meanings in the text that follows. Our conceptions are part of a nascent body of theoretical and clinical literature that explores how transference and countertransference are affected by such cultural factors (e.g., Comas-Diaz & Jacobsen, 1995; Lijtmaer, 2000; Napoli, 1999; Tang & Gardner, 1999). It should be noted, however, that the concept and definition of the working alliance appear to us to be essentially the same whether or not RSM status is part of the equation. At the same time, the issues involved in forming the alliance, as well as impediments to the alliance, may be quite different.

So that we may clarify how traditional conceptions of transference and countertransference are altered when RSM status becomes part of the process, it may be worth reiterating our earlier definitions of these two constructs. Following Gelso and Hayes (1998), we defined transference as the client's experience of the therapist that is shaped by the client's own psychological structures and past and that involves displacement onto the therapist of feelings, attitudes, and behaviors belonging rightfully in earlier significant relationships. Along the same lines, we defined countertransference as the therapist's transference to the client's material, both transference and nontransference communications presented by the client. These definitions have essentially nothing to do with culture (the umbrella term we use for race, ethnicity, and sexual orientation, the complexities of "culture" notwithstanding) (see Phinney, 1996). At the same time, cultural factors certainly may enter into transference and countertransference as so defined. Such cultural factors provide some of the content of transference and countertransference reactions, and they continually shape these reactions.

When cultural factors become a part of the transference and/or countertransference, however, these definitions are insufficient. Two additional conceptions of transference and countertransference appear to us to be central, perhaps vital. The first conception is referred to as cultural transference and countertransference. We define cultural transference as the culture-related distortions of or rigid interpersonal behaviors in response to the therapist that are rooted in the patient's experiences, direct or vicarious, with members of the therapist's RSM group. Likewise, countertransference may be seen as the therapist's culture-related distortions of the patient or rigid interpersonal behaviors rooted in his or her direct or vicarious experiences with members of the patient's RSM group.

To understand the concept of cultural transference (and countertransference), it is important to discuss a characteristic of all transference reactions. Because of their roots in conflict and trauma, transferences have a quality of tenacity—they are not easily given up. Thus, transference and countertransference reactions are more than simple misperceptions of or incorrect generalizations about the therapist based on figures in the patient's past. If this were the case, they could be readily corrected. Instead, transference represents misperceptions that resist correction in the face of disconfirming evidence, and these may be changed only gradually. Transferences are thought to be maintained in the face of contrary data because they protect the patient's self-esteem, avoid anxiety, and provide a way of experiencing others that is familiar and predictable for the patient (for a discussion of the bases for this tenacity, see Gelso & Hayes, 1998). Following this line of thinking, cultural transferences, like transferences that are not culturally based, are maintained with tenacity and need to be treated with the same sensitivity and understanding as noncultural transferences.

This concept of cultural transference has crept into the psychotherapy literature in recent years, as an increasing num-

ber of theoreticians have concerned themselves with cultural issues in psychotherapy. For example, Helms and Cook (1999) offered the following observation.

In addition to avoiding the imposition of her or his racial or cultural biases on the client, the therapist must always be aware of herself or himself as a potential transference symbol for the client. After all, based on the therapist's appearance, the client will decide to which racial group the therapist belongs. More than likely, the client will also make assumptions about other characteristics of the therapist (including socioeconomic status, racial attitudes, cultural traditions) based on the therapist's "racial" appearance. . . . Therefore, the therapist should not ignore the client's perceptions of the therapist's race and culture as aspects of the transferential relationship. It may be difficult for the client to develop transference related to parents (however "parents" are defined) if racial transference develops as an overriding concern. Under such circumstances, the therapist may come to symbolize whatever past traumatic experiences or socialization—personal or vicarious—that the client or the client's salient identity groups have had with members of the therapist's racial or cultural group(s). *Sometimes it is necessary to work through this racial-cultural transference before the more mundane issue of parental transference can even be expected to occur.* (pp. 137–138; emphasis added)

Although we do not see the issue of parental transference as"mundane" or more mundane than cultural transferences, we do share Helms and Cook's (1999) conception that cultural transferences (and countertransferences) need to be carefully attended to, and that they often must be addressed before noncultural transferences, especially if these cultural transferences are negative. We suspect that a significant proportion of the drop-out rate that has been pointed to repeatedly in cross-cultural therapy dyads occurs because cultural transference is ignored or simply not noticed (see the discussion by LaRoche, 1999).

Helms and Cook (1999) were addressing a specific version of cultural transference when they saw the therapist as symbolizing past traumatic experiences or socialization that the client or the client's identity group had encountered with the therapist's identity group. What they pointed out is a negative transference that specifically derives from the patient's expectations of the therapist, based on the therapist's cultural group. The experience of oppression seems key in their definition (e.g., their focus on cultural traumatization). Similarly, transference for LaRoche (1999) is not simply a repetition of the patient's relationship with parental figures; transference reactions are also about "the various sociocultural meanings regarding therapist's authority and assumed group membership" (p. 391). However, cultural transferences may be negative or positive, and although they may reflect traumatization by the therapist's identity group, they may also

represent a wide range of projections onto the therapist. For example, as LaRoche noted, "Latinos have favorable opinions of Americans and expect much from psychotherapy. These cultural expectations need to be explored and addressed as soon as possible in order to prevent disappointment" (p. 391).

It is important to recognize that cultural transferences also occur when the patient is a member of the majority group and the therapist is a member of a minority group. Cultural transferential experiences of fear, disdain, superiority, and comfort may occur when the patient is White and the therapist is Black. Thompson (July 2000, personal communication), for example, notes that some White patients of Black therapists persist in treatment because these patient's feel like they are the "colored person" in their families of origin. Thus, they identify with the therapist's outsider status and feel as though they are with a therapist who can understand them. The Black therapist may be experienced as safe and caring as part of such a transference, but they may also be experienced as the devalued "servant." When gender is added to the racial/ethnic mix, complex gender × race interactions are likely. For example, the White male patient may project racial aggression onto the Black male therapist, but may transferentially experience the Black female therapist as a nurturing "nanny."

The examples provided by Thompson lead to the second conception of transference and countertransference that appears central when RSM members are involved in treatment. This second conception is referred to as culturally reinforced transference and countertransference. These transferences are similar to cultural transferences, but they have a deeply significant additional feature: they are rooted in early childhood experience. We define these transferences as culture-related distortions of or rigid interpersonal behaviors in response to the therapist that are connected to and partly stemming from unresolved conflicts early in the patient's life with significant others. (The definition of countertransference is parallel to this, with the only difference being that it is a feature of the therapist rather than the patient.) In such transferences, the cultural component is fueled partly by the earlier roots and partly by cultural phenomena. In this way, the culturally reinforced transference is a combination of cultural and noncultural transference. An example of culturally reinforced transference would be the Black male patient who fears that his White male therapist will try to make him subservient in their relationship, as White people historically have done with Black people; this patient projects these impulses onto his therapist. The tendency to experience the therapist as diminishing (the cultural aspect of the transference) is fueled by the patient's early and ongoing experience with his father, an anxious man who, out of his own insecurity, diminished each of his four children, fearing that they would surpass him in their achievements. Another example is the following case treated by the first author.

The client was a 28-year-old gay male who happened to come from the same small town as the therapist, a town whose inhabitants tended to have very conservative attitudes toward sexuality and most other political issues. The therapist was an experienced psychodynamically oriented practitioner who was male and heterosexual. The patient harbored a fear that the therapist, as a heterosexual male, was contemptuous of the patient, partly for his sexual orientation (cultural transference). As part of this projection, the patient feared and felt that the therapist was critical of him for not being effective at mechanical things (e.g., repairing his car, fixing broken faucets). In fact, the therapist himself was not mechanical and was comfortable with this fact. The patient's father, however, was a plumber by trade and was abusively critical of his children for all of their "shortcomings" including their not being handy around the house (noncultural transference). Thus, the cultural transference around conventional masculinity was fueled by more primitive issues revolving around the patient's relationship with an abusive father.

Because culturally reinforced transference is rooted in the patient's direct and vicarious cultural experience as well as in unresolved issues (e.g., with primary caretakers), we believe them to be among the most complex and difficult transferences for the patient to understand and work through. The basis for the projections may be seen as double-barreled.

We have been discussing three kinds of transference: cultural, culturally reinforced, and noncultural. A word should be added about so-called noncultural transferences. Surely cultural factors enter into all transferences and countertransferences. From our perspective, cultural factors cannot *not* come into play. And yet cultural factors revolving around race/ethnicity and sexual orientation are not fundamentally important in all transferences, even though all humans are partly cultural beings. From our perspective, noncultural transferences are fueled mostly by individuals' unique histories of interactions with significant others rather than by their contact with shared cultural realities.

It should also be noted that there is a fourth category with respect to the therapeutic relationship and the participants' reactions to one another: nontransference. Not all client perceptions, to say nothing about therapist perceptions, are transferential. Using the patient as an example, much of the patient's experience and perceptions of the therapist are rooted in reality, and it is important that this be understood and acknowledged. Furthermore, even transferential reactions nearly always have a realistic element and in this sense are coconstructed by therapist and client. Because of this coconstruction, it is helpful for the therapist to consider his or her contribution to the transference. Doing so provides the patient with validation, while at the same time fostering the discovery of how the patient reenacts his or her issues in the transference. In any event, it is especially important that the therapist maintain a high level of self-awareness during cross-cultural therapy. As therapists, we are surely not immune to culturally based distortions (cultural countertransferences); in fact, we suspect that the intimacy and intensity

Table 1. *A 2 × 2 Categorization of Transferences Based on Cultural and Familial Causal Factors*

	Familial factors	
	Present	Absent
Cultural Factors		
Present	Culturally Reinforced Transference	Cultural Transference
Absent	Noncultural Transference	Nontransference

of the therapeutic relationship pulls for cultural counter-transferences. So it behooves the culturally sensitive therapist to stay open and curious about whether his or her patient's reactions represent cultural or culturally reinforced transference projections or whether they are realistic and accurate perceptions of the therapist, picking up on subtle aspects of the therapist's unconscious cultural prejudices, idealizations, and other unresolved issues.

Table 1 presents the transferences we have been discussing. As can be seen, a 2 × 2 categorization is possible based on the presence or absence of cultural or familial factors as fundamental causal agents. The result is the four clusters we have explored: cultural transference, culturally reinforced transference, noncultural transference, and nontransference.

We have discussed the four clusters as if they represented nominal categories. This has been for the sake of convenience and communication. From a scientific perspective, however, it may be more useful to think of two dimensions: a cultural and a noncultural dimension. Thus, to one degree or another, transference and countertransference may be seen as residing on these two dimensions. Such an ordinal system permits scientific measurement and analysis and may better represent the "true" state of the world as regards these phenomena.

Let us now explore working alliance, transference, and countertransference in psychotherapy dyads in which one or more participant is a member of an RSM group. We begin by examining the working alliance and how it is affected by RSM factors. Note that in the following discussion we incorporate the comments of the experts we interviewed. When we do so, their comments are cited as personal communications.

The Working Alliance, Race/Ethnicity, and Sexual Orientation

Based on clinical experience as well as research findings over many years about how racial pairings influence therapy process and outcomes (see, e.g., reviews by Coleman, Wampold, & Casili, 1995; Sue, Zane, & Young, 1994), one can readily offer the following general proposition.

Proposition 5: Perceived RSM similarity aids the initial formation of a strong working alliance, whereas RSM dissimilarity makes initial alliance development more complex

and difficult. Note that the focus here is on perceived rather than actual similarity. It should also be noted that there are many instances in which actual or perceived similarity is either irrelevant or even detrimental to the formation of working alliance. We have more to say about these situations subsequently.

Regarding the benefits of similarity, Cook (personal communication, June 2000) cites a Black blues saxophonist who commented to his appreciative Black listeners that he experienced an "abundance of understanding" from them. To Cook, this is an apt metaphor for working alliance formation when participants from the same RSM group form the therapeutic dyad. In other words, an immediate alliance may be formed in which the client expects to be understood, and the therapist is especially ready to understand. We suspect that this early alliance formation (or this expectation and readiness for an abundance of understanding) occurs more broadly when members of sociocultural groups that have been the recipients of severe societal oppression form the therapeutic dyad.

On the other side of the ledger, when the members of the dyad are of different races or sexual orientations, the working alliance may take special efforts to cultivate. This is especially the case when one participant is a majority group member and the other is a minority group member (for both race/ethnicity and sexual orientation). This is so because the initial cultural and culturally reinforced transferences are likely to be negative, revolving around fear and an expectation of nonunderstanding. We have more to say about such transferences in the text that follows, but suffice it to say that such transferences create an initial stumbling block for healthy alliance formation. It should be noted that initial cultural and culturally reinforced countertransferences are often also at work and serve as impediments to sound alliance formation. The possibility (perhaps likelihood) of cultural countertransferences points to the fact that not all cultural expectations are transferential. At times (some would say usually), the client's negative expectations of the therapist is on target. The therapist does, in fact, harbor conscious or unconscious conflicts around race and/or sexual orientation.

Although the initial alliance formation is likely to be positive when therapy participants are of the same race or sexual orientation, a word of caution is in order. Rios (personal communication, July 2000) suggests that a superficial alliance based on race/ethnicity similarity can jeopardize the working through of transference issues. This superficial alliance may also be thought of as a premature alliance or even a pseudo-alliance. In such cases, the therapist seeks to cultivate an alliance too quickly based on race/ethnicity (or sexual orientation, we would add). Rios provides a case example that clearly illuminates this issue.

The client was a adolescent Black male. The therapist was Latina who appeared White. The client presented himself as very angry and was quite challenging and questioning of the therapist in their initial meeting. Thus, he entered

therapy with a psychological base of anger that likely had components related to race and gender. The therapist "corrected" the client's assumption that the therapist couldn't understand his experience as a person of color by telling him that she herself was a Person of Color. She did this because of her countertransference-based need to diffuse the client's anger; but, in doing so, she lost an opportunity to explore affectively his fears about surviving in a White world. Rather than acting out her countertransference in this way, the therapist wished she had built an alliance with the client to the point where he could have really dealt in a deep way with his racial issues. Instead she went for a "quick fix" for the strain in the alliance by "coming out" as Latina.

Rios also notes that, had the therapist been White, her defensive reaction to the client's challenge may have been on the order of "Yes, I could never really understand your experience because People of Color have such a hard time of it." In other words, the therapist's cultural countertransference would likely have been around racial guilt. In these examples, the therapist's fear of potentially alliance-damaging transference created a countertransference reaction that, in turn, sought to create a quick alliance. Such premature alliances, in keeping with Rios' observations, are unlikely to be sound and stable.

Within-Group Factors in Alliance Formation

The hypothesis offered in the last section may be termed the demographic similarity hypothesis. To our mind, both clinical experience and empirical evidence support this hypothesis. However, we do not believe this main effect hypothesis goes far enough, and in certain instances it may, in fact, be misleading. There are two reasons for our belief in the insufficiency of the demographic similarity hypothesis. First, in its focus on demographics, the hypothesis does not address the underlying psychological factors that more directly affect alliance. For example, in cross-racial therapy dyads, within-client and within-therapist constructs, such as racial trust–mistrust (Terrell & Terrell, 1984; Thompson, Worthington, & Atkinson, 1994) and perceived attitude and value similarity (Atkinson & Lowe, 1995), have been empirically demonstrated to affect preferences for therapist. In fact, in a series of laboratory analogues, Atkinson and his collaborators (see Atkinson & Lowe, 1995) demonstrated that most racial/ethnic minority participants rank factors such as attitude, value, and personality similarity as more important in determining their preferences for a counselor than they do racial similarity.

The empirically supported importance of internal, psychological processes (e.g., racial attitudes and values) within the client and therapist in RSM dyads hints at the second reason why the demographic similarity hypothesis does not go far enough. That is, these within-group factors are likely to moderate the effects of RSM pairings on the formation of working alliance. Whereas the demographic similarity hypothesis is a main effect hypothesis, the within-group moderating hypothesis is an interaction hypothesis in the statistical sense. And, as is the case with interaction effects generally, they qualify main effects. In other words, interactions ordinarily inform us about the conditions under which main effects may and may not be operative. In terms of working alliance formation in the RSM pairings of interest, we suggest the following hypothesis: *In therapy dyads in which one or more participant is a racial or sexual orientation minority, within-therapist and within-patient factors will significantly influence or moderate the effect of perceived race and sexual orientation similarity on working alliance formation.*

Innumerable examples may be provided of the within-group moderating hypothesis in the area of RSM pairings and the working alliance. For example, White therapists may have difficulty establishing a productive alliance with Black patients who have high levels of cultural or racial mistrust, regardless of the therapists' level of self-awareness, knowledge, and experience concerning race. On the other hand, knowledgeable and self-aware White therapists may have a relatively easy time forming an alliance with Black patients who have a lower level of cultural or racial mistrust. Drescher (May 2000, personal communication) offers an illuminating example of when similarity impedes rather than fosters the alliance.

A patient who was confused about his sexual orientation chose a gay therapist with the unconscious aim of (a) externalizing his internalized homophobia by denigrating the therapist and (b) colluding with the denigrated therapist in maintaining an antigay stance. The formation of a strong working alliance was a delicate, arduous, and complex matter in this case. The countertransference difficulties that such a case inevitably creates further complicated the development of a sound alliance.

Examples such as these may be usefully conceptualized from the perspective of the culturally based identity theories that have emerged into prominence in recent years. Such theories have been developed in the areas of race/ethnicity (e.g., Helms, 1995; Rowe, Behrens, & Leach, 1995) and sexual orientation (e.g., Fassinger & Miller, 1996; McCarn & Fassinger, 1996). In discussing racial identity models, Helms and Cook (1999) stated that:

In actuality, racial (in fact, "socioracial") identity models are psychological models because they intend to explain individuals' intrapsychic and interpersonal reactions to societal racism in its various manifestations. That is, they are descriptions of hypothetical intrapsychic pathways for overcoming internalized racism and achieving a healthy socioracial self-conception under varying conditions of racial oppression. We find these models useful for assess-

ing the influence of racial factors on the client's concerns as well as the reactions of the client and therapist to one another. (p. 81).

Perhaps the most comprehensive, integrative, and heuristic identity theory is the racial identity theory of Helms (1984, 1990, 1995; Helms & Cook, 1999). Because this theory has considerable value for understanding therapeutic alliance formation and subsumes many of the specifics we have noted here, we provide a brief summary. In its most recent rendition, Helms' racial identity theory conceptualizes racial identity development in terms of ego identity statuses, which in turn are defined as "cognitive-affective-conative intrapsychic principles for responding to racial stimuli in one's internal and external environments" (Helms & Cook, p. 84). Models are offered by Helms for People of Color (i.e., Asian, African, Latino/Latina, and Native Americans living in the United States) and for White people. In the model for people of color, there are six ego identity statuses, and these range from the Conformity status to the Integrative Awareness status. At the lowest level (Conformity), the person devalues his or her own race and is oblivious to social racial concerns. At the highest level (Integrative Awareness), the person values his or her own collective identity and empathizes with members of other oppressed groups. Life decisions are motivated by globally humanistic self-expression. In the four statuses in between the highest and lowest, individuals range from being ambivalent and confused about their socioracial group commitment, to idealizing their own racial group while denigrating the majority group, to being deeply immersed in one's own group, to moving toward acceptance of one's group and an ability to respond objectively to members of the dominant group.

For the White racial identity model, there are seven ego identity statuses. At the least developed status (Contact), the person is satisfied with the racial status quo and, in fact, is oblivious to racism and to his/her participation in it. At the most developed status (Autonomy), the person evidences "informed positive socioracial-group commitment, use of internal standards for self-definition, capacity to relinquish the privileges of racism" (Helms & Cook, 1999, p. 91). In the five statuses in between, the person may range from great ambivalence and anxiety about racial dilemmas, to idealization of Whiteness and denigration of other groups, to intellectualized racial openness and subtle White superiority, to a search for the personal meaning of whiteness and racism.

As can be seen from this brief glimpse of Helms' racial identity theory for both White people and People of Color, the racial identity status of both client and therapist has major implications for the quality of working alliance that is formed. Certain combinations would make for a ready alliance formation, whereas others would make it nearly impossible for a White person and a Person of Color (whomever is in the therapist or client role) to form a workable alliance (e.g., see Helms & Cook, 1999). These types of dynamics are

also relevant for ethnic identity (as differentiated from racial identity), where alliance formation may be significantly affected by the similarities and differences in clients' and therapists' culture-based values, worldviews, and communication styles (Chin et al., 1993; La Roche, 1999).

The emerging gay and lesbian identity theories have a similar thrust to racial identity theories in the focus on awareness of and response to societal oppression, and these theories also have similar implications for alliance formation. One point of divergence from racial identity theories, however, is related to a fundamental difference between sexual orientation and race/ethnicity. Whereas racial/ethnic categories are largely based on one's visible characteristics (e.g., skin color) and characteristics of one's family of origin (e.g., language, religion), sexual orientation categories are largely based on a factor that is both invisible and open to change: romantic and sexual attractions. Thus, a unique task of sexual orientation identity development is discovering (or perhaps rediscovering) who one is in terms of one's attractions. Because of this component of identity development, Fassinger and her colleagues proposed that lesbian and gay identity formation consists of two interrelated processes (Fassinger & Miller, 1996; McCarn & Fassinger, 1996). One process involves individual sexual awareness and identification, and the other process involves reference group identification. Although few parallels to White racial identity theory have been offered in the sexual orientation literature, models of heterosexual identity have been proposed recently (Worthington & Mohr, in press).

Many specific hypotheses are possible for working alliance formation with RSM pairings based on racial and sexual orientation identity theories. Although the statement of such hypotheses is beyond the scope of the present article, this is a highly fruitful line of inquiry for future theoretical and empirical work.

In the sections that follow, we discuss transference, countertransference, and the relation of these to working alliance in RSM pairings. As we examine these constructs, our discussion of the within-group moderating hypothesis and of racial and sexual orientation identity as moderating variables should be kept in mind.

Transference and RSM Pairings

Issues of authority are inherent in the client–therapist relationship as well as in relationships between members of minority and majority cultures. Thus, as Drescher (personal communication, May 2000) points out, transferential issues related to authority are especially likely to occur when the client is a minority culture member and perceives the therapist to be a majority culture member. In such cases, the client is likely to believe that the therapist holds majority views (e.g., is homophobic, is ignorant regarding gay and lesbian issues). Although these beliefs reflect cultural transferences, they also likely reflect some degree of reality. To be effective, the therapist must be comfortable exploring and possibly acknowledging his or her own issues related to the client's pro-

jection or perception. Once again, we point to the issue of cul-
tural countertransference and its importance when working
with members of racial or sexual orientation groups that are
different from the therapist's.

As a case example of these issues, Drescher describes a
gay patient who led a very active sexual life and was con-
vinced that the therapist disapproved. The patient may be on
target in that there may be a realistic basis to his beliefs about
the therapist's disapproval, and the therapist may be out of
touch with his or her own disapproval (as well as the value
and countertransference material that may be associated with
this disapproval). On the other hand, the client's belief may
reflect transferential distortions related to the client's gay
identity. Such forms of cultural and culturally reinforced trans-
ference may be most common in clients who are in the earli-
er stages of identity formation, wherein negative views are
held either of their own sexual orientation or of heterosexu-
ality. Thus, the therapist may represent an amalgam of roles,
such as parent and representative and enforcer of oppressive
social norms.

Cultural and culturally reinforced transferences also figure
prominently into the relationship when the therapist is from
a minority culture. Thompson (personal communication, July
2000) notes that being a Black therapist is a lightening rod
for transference issues. We would broaden the conception
and add that, when one or more participant in the dyad is a
member of a historically and/or currently oppressed group,
the therapist's race or sexual orientation is a lightening rod
for transference. Sometimes these transferences are subtle, at
times becoming manifest in the form of dreams. Thompson,
for example, describes a patient whose dreams are related to
what we are calling cultural and culturally reinforced trans-
ference:

> In one client's dream, the White client showed the thera-
> pist a house the therapist clearly could not afford. This
> tapped into a real issue for the therapist, but it also re-
> flected a transferential one involving the assertion of soci-
> etal norms around race and class. In another dream, the
> client was on a sinking boat. The client shared that the ther-
> apist was the captain, but then showed some uncertainty
> about this: "Well, maybe it was you, but it definitely was
> a Black person." The therapist's countertransference reac-
> tion to this was shame—shame that was based on societal
> oppression (i.e., the therapist as herself or as any other
> Black person was not good enough to keep the boat afloat).
> Analysis of the dream with the client also suggested that
> the therapist was indistinguishable from other Black peo-
> ple, that all Black people are alike.

As Thompson notes, White clients are often anxious about
sharing race-related transference material due to fears of be-
ing seen as racist and jeopardizing the working alliance with
racial minority therapists. In such cases, the working alliance
itself can serve as a buffer against such anxieties and allow

for a relationship in which it feels safe to explore cultural and
culturally reinforced transferences.

As is its nature, transference reactions may be highly com-
plex; thus complexity is magnified when therapist and client
are from different RSM groups. Drescher (personal commu-
nication, May 2000) describes a case in which a man who
was confused about his sexual orientation chose a gay thera-
pist. This choice appeared to be aimed at the client's uncon-
sciously externalizing his homophobia by denigrating the
therapist and colluding with the denigrated therapist in main-
taining an antigay stance. These unconscious maneuvers
were also viewed by the therapist as a "transference test" or
a "transference learning experience" in that the patient want-
ed to see how the therapist would deal with homophobia.
Would the therapist accept the antigay stance? Would the
therapist reject himself and the client? How would the ther-
apist withstand homophobic assaults?

Although we believe that awareness of cultural and cul-
turally reinforced transference enriches therapists' work with
clients, it is critical to attend to alternative interpretations of
client behavior that may initially appear transferential. Ther-
apists who are unfamiliar with the normative behaviors of
certain cultural groups may mistakenly believe that clients
are engaging in transference when they are actually acting in
accord with culturally prescribed roles. For example, clients
from cultures that emphasize deference for medical profes-
sionals may exhibit behavior that appears unusually passive
or submissive to some therapists. Such therapists may auto-
matically interpret this behavior as a sign of distortion or
pathology rather than an expression of social norms.

Countertransference Management
as a Necessary Ingredient

Earlier in this article, we addressed the critical importance
of therapists coming to grips with their countertransference
reactions. We noted that such reactions are inevitable, and
that they could be for better or worse, depending on how they
are dealt with or managed. Based on the burgeoning litera-
ture on multicultural therapy, our own personal experiences
as therapists, and the four interviews we have reported, it
seems inarguable that the management of cultural and cul-
turally reinforced countertransference is also a key ingredi-
ent of successful treatment when one or more participant in
the dyad is members of a RSM group. In fact, we would sug-
gest that it is the most potent ingredient. This is because in
our society race, ethnicity, and sexual orientation are such
emotionally provocative factors and because long-standing
issues of oppression inevitably enter into the therapeutic re-
lationship. The therapist cannot be immune to these issues,
and they must affect his or her emotional reactions in a range
of ways. Ongoing self-monitoring and the willingness to face
one's fears, conflicts, and idealizations about race and sexu-
al orientation are of fundamental significance for therapists.

Within psychoanalysis, the topic of countertransference
had the features of a taboo for many years; it is only recent-

ly that anything like an open discussion of the topic has oc- curred. Few personal examples are offered in the literature. When race and sexual orientation are part of treatment, even fewer examples are provided. Rios (personal communica- tion, July 2000) sidesteps such prohibition and provides the following case vignette from her work.

> The client was a Colombian woman who suffered from panic attacks, and the therapist's ethnic background was also Colombian. Racially, both were White. The client clearly could have benefitted from desensitization therapy, and the therapist ordinarily referred to other therapists for such work. In this case, however, the therapist conducted the treatment herself. On reflection after termination, the therapist believed that she treated the case because of her identification with the client as a Colombian. The therapist also believed she may not have been doing enough ana- lytic work (her preferred orientation) about how the panic came to be. She felt this avoidance of analytic work was due to the fact that the panic likely was rooted in the wide- spread violence and terror in Colombia, and this would be painful for both client and therapist to explore. Thus, the focus on behavioral work may have reflected a posttrau- matic avoidance on the part of both therapist and client.

In this case, it is clear that countertransference can be positive as well as negative in valence, and that positive countertrans- ference can work against constructive change, just as negative countertransference can if it is not understood and managed in one way or another. Helping clients avoid painful issues around culture, we would add, may be as harmful as therapists acting out of negative countertransference feelings.

Interplay of Alliance and Transference/ Countertransference in RSM Pairings

We have offered a number of suggestions about the recip- rocal impact of working alliance and transference/counter- transference in general. These hypothesized relationships, as noted earlier, were expected to be maintained in RSM pair- ings. Naturally, though, the content (i.e., the participants' feelings, attitudes, expectations, and reactions) of the al- liance and transference/countertransference would be differ- ent and would reflect cultural issues the a greater extent. We provide a few examples of these interrelationships.

As an example of how an initial negative cultural trans- ference (or perhaps culturally reinforced transference) may preclude the development of a sound working alliance, Cook (personal communication, June 2000) describes a White client being in an intake interview with a Black therapist. The client explored his issues deeply during the session and shared a great deal of feelings. This included emotion-laden discussion of earlier suicide attempts. By the end of the in- take interview, a sense of closeness had been established in the therapist's opinion. However, the client revealed that she could not imagine being seen by a Black therapist. In this

case, the cultural transference issues were simply too strong and negative, and they poisoned the emerging relationship. In such a case, the therapist was helpless to deal with the transference in a way that allowed therapy to continue.

When the alliance does develop, it can facilitate the emer- gence and resolution of transference and can helpfully affect countertransference. This facilitating effect is exemplified by Cook in the following vignette.

> A White male client was in therapy for 3 months with a very light-skinned Black therapist. A relatively strong al- liance had developed over this period. In a particular ses- sion during the fourth month the client noticed for the first time that most of the therapist's photos were of Black friends and family. The client was surprised by this and subsequently discovered that his therapist was Black. This discovery led to a more complex transference than had been the case. The client retained the eroticized transfer- ence that had developed prior to his discovery, but could no longer maintain the positive transference based on the assumed Whiteness of the therapist. Instead, he developed a negative cultural transference around his beliefs that a Black person cannot be a doctor. Furthermore, he had to struggle with the realization that his erotic reactions were now directed toward a Black person. The previously es- tablished working alliance was important in negotiating this change in the relationship, although the negative cul- tural transference, at the same time, did adversely affect the alliance, creating a rupture of sorts. Importantly, the positive working alliance also seemed to help the therapist avoid negative countertransference related to the client ini- tially viewing her as White, given her own issues around appearing too White.

When cultural and culturally reinforced transference is- sues invade the working alliance and pose a threat to the over- all therapeutic relationship, it is important that therapists bring these into the open and help clients explore them. Just as in therapy in general, the need for exploration is especial- ly great when transferences are negative.

Concluding Comment

In this article we have explored the interrelations of working alliance, on the one hand, and both transference and coun- tertransference, on the other. We extended our theoretical dis- cussion of these interrelations to the psychotherapy situation in which at least one member of the treatment dyad is a mem- ber of a racial/ethnic or sexual orientation minority.

Although the empirical and theoretical literature on diver- sity topics in therapy has burgeoned in recent years, writing on psychoanalytic constructs has trailed behind. Yet there is a nascent body of literature on race/ethnicity and sexual ori- entation factors in the therapy relationship from psychody- namic and psychoanalytic perspectives (Drescher, 1998; Jack-

son & Greene, 2000; Leary, 1995; Roland, 1996; Thompson, 1995), and our conceptions are best viewed as part of these developments. Our hope is that the theoretical and clinical reflections that we have made will help generate further em-

pirical inquiry and theory development on the topic of working alliance, transference, and countertransference in psychotherapy (psychoanalytic or otherwise) in which factors related to race and sexual orientation are central.

REFERENCES

Adelstein, D., Gelso, C., Haws, J., Reed, K., & Spiegel, S. (1983). The change process following time-limited therapy. In C. Gelso & D. Johnson (Eds.), *Exploratons in time-limited counseling and psychotherapy* (pp. 63–81). New York: Columbia University, Teachers College Press.

Andersen, S. M., & Berk, M. S. (1998). The social-cognitive model of transference: Experiencing past relationships in the present. *Current Directions in Psychological Science, 7,* 109–115.

Atkinson, D. R., & Lowe, S. M. (1995). The role of ethnicity, cultural knowledge, and conventional techniques in counseling and psychotherapy. In J. Pontorotto, J. M. Casas, & L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 387–414). Thousand Oaks, CA: Sage.

Barber, J. P., Foltz, C., & Weinryb, R. M. (1998). The Central Relationship Questionnaire: Initial report. *Journal of Counseling Psychology, 45,* 131–142.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, research, and practice, 16,* 252–260.

Bowlby, J. (1979). *The making and breaking of affectional bonds.* New York: Routledge & Kegan Paul.

Cahill, S. P., Carrigan, M. H., & Evans, I. M. (1998). The relation between behavior theory and behavior therapy: Challenges and promises. In J. J. Plaud & G. H. Eifert (Eds.), *From behavior theory to behavior therapy.* Boston, MA: Allyn & Bacon.

Chin, J. L., Liem, J. H., Ham, M. D., & Hong, G. K. (1993). *Transference and empathy in Asian American psychotherapy: Cultural values and treatment needs.* Westport, CT: Praeger.

Coleman, H. L. K., Wampold, B. E., & Casili, S. L. (1995). Ethnic minorities' ratings of ethnically similar European American counselors: A meta-analysis. *Journal of Counseling Psychology, 42,* 55–64.

Comas-Diaz, L., & Jacobsen, F. M. (1995). The therapist of color and the White patient dyad: Contradictions and recognitions. *Cultural Diversity and Ethnic Minority Psychology, 1,* 93–106.

Drescher, J. (1998). *Psychoanalytic therapy and the gay man.* Hillsdale, NJ: Analytic Press.

Fassinger, R. E., & Miller, B. A. (1996). Validation of a model of sexual identity development for a sample of gay men. *Journal of Homosexuality, 32,* 53–79.

Freud, S. (1953). Fragment of an analysis of a case of hysteria. In J. Strachey (Ed.), *Standard edition of the complete psychological works of Sigmund Freud* (Vol. 7, pp. 15–122). London: Hogarth Press. (Original work published 1905)

Freud, S. (1958). On the beginning of treatment: Further recommendations on the technique of psychoanalysis. In J. Strachey (Ed.), *Standard edition of the complete psychological works of Sigmund Freud* (pp. 122–144). London: Hogarth Press. (Original work published 1913)

Freud, S. (1959a). Future prospects of psychoanalytic psychothera-

py. In J. Strachey (ed.), *Standard edition of the complete psychological works of Sigmund Freud* (Vol. 20, pp 87–172). London: Hogarth Press. (Original work published 1910)

Freud, S. (1959b). The dynamics of transference. In E. Jones (Ed.) & J. Riviere (Trans.), *Collected papers* (Vol. 2, pp. 312–322). New York: Basic Books (original work published 1912)

Friedman, S. M., & Gelso, C. J. (2000). The development of the Inventory of Countertransference Behavior. *Journal of Clinical Psychology, 56,* 1221–1235.

Fromm-Reichmann, F. (1950). *Principles of intensive psychotherapy.* Chicago, IL: University of Chicago Press.

Fuertes, J. N., & Gelso, C. J. (2000). Hispanic counselors' race and accent and Euro Americans' Universal-diverse orientation: A study of initial perceptions. *Cultural Diversity and Ethnic Minority Psychology, 6,* 211–219.

Gelso, C. J., & Carter, J. A. (1985). The relationship in counseling and psychotherapy: Components, consequences, and theoretical antecedants. *Counseling Psychologist, 13,* 155–243.

Gelso, C. J., & Carter, J. A. (1994). Components of the psychotherapy relationship: Their interaction and unfolding during treatment. *Journal of Counseling Psychology, 41,* 296–306.

Gelso, C. J., Fassinger, R. E., Gomez, M. J., & Latts, M. G. (1995). Countertransference reactions to lesbian clients: The role of homophobia, counselor gender, and countertransference management, *Journal of Counseling Psychology, 42,* 356–364.

Gelso, C. J., & Hayes, J. A. (1998). *The psychotherapy relationship: Theory, research, and practice.* New York: John Wiley.

Gelso, C. J., & Hayes, J. A. (in press). The management of countertransference. In J. Norcross (Ed.), *A guide to psychotherapy relationships that work.* Oxford: Oxford University Press.

Gelso, C. J., Hill, C. E., & Kivlighan, D. M. (1991). Transference, insight, and the counselor's attention during a counseling hour. *Journal of Counseling and Development, 69,* 428–433.

Gelso, C. J., Hill, C. E., Mohr, J. J., Rochlen, A. B., & Zack, J. (1999). Describing the face of transference: Psychodynamic therapists' recollections about transference in cases of successful long-term therapy. *Journal of Counseling Psychology, 46,* 257–267.

Gelso, C. J., & Johnson, D. H. (1983). *Explorations in time-limited counseling and psychotherapy.* New York: Columbia University, Teachers College Press.

Gelso, C. J., Kivlighan, D. M., Wine, B., Jones, A., & Friedman, S. C. (1997). Transference, insight, and the course of time-limited therapy. *Journal of Counseling Psychology, 44,* 209–217.

Gelso, C. J., Latts, M. G., Gomez, M. J., & Fassinger, R. E. (in press). Countertransference management and therapy outcome: An initial evaluation. *Journal of Clinical Psychology.*

Graff, H., & Luborsky, L. L. (1977). Long-term trends in transference and resistance: A report on a qualitative-analytic method applied to four cases of psychoanalyses. *Journal of the American Psychoanalytic Association, 25,* 471–490.

Greenson, R. R. (1965). The working alliance and the transference neurosis. *Psychoanalytic Quarterly, 34,* 155–181.

Greenson, R. R. (1967). *The technique and practice of psychoanalysis* (Vol. 1). New York: International Unviersities Press.

Hayes, J. A., McCracken, J. E., McClanahan, M. K., Hill, C. E., Harp, J. S., & Carozzoni, P. (1998). Therapist perspectives on countertransference: Qualitative data in search of a theory. *Journal of Counseling Psychology, 45,* 468–482.

Hayes, J. A., Riker, J. R., & Ingram, K. M. (1997). Countertransference behavior and management in brief counseling: A field study. *Psychotherapy Research, 7,* 145–153.

Helms, J. E. (1984). Toward a theoretical explanation of the effects of race on counseling. *The Counseling Psychologist, 12,* 153–165.

Helms, J. E. (Ed.). (1990). *Balck and white racial identity: Theory, reserach, and practice.* Westport, CT: Greenwood.

Helms, J. E. (1995). An update of Helms' White and People of Color racial identity models. In J. G. Pontorotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 181–198). Thousand Oaks, CA: Sage.

Helms, J. E., & Cook, D. A. (1999). *Using race and culture in counseling and psychotherapy: Theory and process.* Boston: Allyn & Bacon.

Jackson, L. C., & Greene, B. (Eds.). (2000). Psychotherapy with African-American women: Innovations in psychodynamic perspectives and practice. New York: Guilford Press.

Kohut, H. (1977). *The restoration of the self.* New York: International Universities Press.

Kohut, H. (1984). *How does analysis cure?* Chicago, IL: University of Chicago Press.

Lambert, M. J., & Bergin, A. E. (1994). The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed., pp. 143–189). New York: John Wiley.

LaRoche, M. J. (1999). Culture, transference, and countertransference among Latinos. *Psychotherapy: Theory/Research/Practice/Training, 36,* 389–397.

Leary, K. 1995). "Interpreting in the dark": Race and ethnicity in psychoanalytic psychotherapy. *Psychoanalytic Psychology, 12,* 127–140.

Ligiero, D. P., & Gelso, C. J. (2001). The counselor and the counseling relationship: Countertransference, attachment, and the working alliance. Paper presented at the 2000 Annual Convention of the American Psychological Association, Washington, DC.

Lijtmaer, R. M. (2000). Psychotherapy with multicultural populations. *Issues in Psychoanalytic Psychology, 22,* 35–42.

Luborsky, L. L., & Barber, J. P. (1994). Perspectives on seven transference-related measures applied to the interview with Ms. Smithfield. *Psychotherapy Research, 3/4,* 172–183.

Luborsky, L. L., & Crits-Christoph, P. (1990). *Understanding transferences.* New York: Basic Books.

Luborsky, L. L., Popp, C., Luborsky, E., & Mark, D. (1994). The core conflictual relationship theme. *Psychotherpy Research, 4,* 172–183.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68,* 438–450.

Miville, M. L., Gelso, C. J., Pannu, R., Liu, W., Touradji, P., Hol-

loway, P., & Fuertes, J. (1999). Appreciating similarities and valuing differences: The Miville-Guzman Universality–Diversity Scale. *Journal of Counseling Psychology, 46,* 291–307.

McCarn, S. R., & Fassinger, R. E. (1996). Re-visioning sexual minority identity formation: A new model of lesbian identity and its implications for counseling and research. *Counseling Psychologist, 24,* 508–534.

Mohr, J. J., & Fassinger, R. E. (2000). Measuring dimensions of lesbian and gay male experience. *Measurement and Evaluation in Counseling and Development, 33,* 66–90.

Mohr, J. J., & Rochlen, A. B. (1999). Measuring attitudes regarding bisexuality in lesbian, gay male, heterosexual populations. *Journal of Counseling Psychology, 46,* 353–369.

Morrow, S. L. (2000). First do not harm: Therapist issues in psychotherapy with lesbian, gay, and bisexual clients. In R. M. Perez, K. A. DeBord, & K. J. Bieshke (Eds.), *Handbook of counseling and psychotherapy with lesbian, gay, and bisexual clients* (pp. 137–156). Washington, DC: American Psychological Association.

Napoli, M. (1999). The non-Indian therapist working with American Indian clients: Transference-countertransference implications. *Psychoanalytic Social Work, 6,* 25–47.

Orange, D., Atwood, G., & Stolorow, R. (1997). *Working intersubjectivity: Contextualism in psychoanalytic practice.* Hillsdale, NJ: Analytic Press.

Patton, M. J., Kivlighan, D. M., & Multon, K. D. (1997). The Missouri psychoanalytic counseling research project: Relation of changes in counseling process to client outcomes. *Journal of Counseling Psychology, 44,* 189–208.

Peabody, S. A., & Gelso, C. J. (1982). Countertransference and empathy: The complex relationship between two divergent concepts in counseling. *Journal of Counseling Psychology, 29,* 240–245.

Phinney, J. S. (1996). When we talk about American ethnic groups, what do we mean? *American Psychologist, 51,* 918–927.

Piper, W. E., Ogrodniczuk, J. S., Joyce, A. S., McCallum, M., Rosie, J. S., & O'Kelly, J. G. (1999). Prediction of dropping out in time-limited, interpretive individual psychotherapy. *Psychotherapy: Theory/Research/Practice/Training, 36,* 144–122.

Pope-Davis, D. B., & Coleman, H. C. (Eds.). (1997). *Multicultural counseling competencies: Assessment, education and training, and supervision.* Thousand Oaks, CA: Sage.

Rhodes, R. H., Hill, C. E., Thompson, B. J., & Elliott, R. (1994). Client retrospective recall of resolved and unresolved misunderstanding events. *Journal of Counseling Psycholoyg, 31,* 473–483.

Rios, P., Corbett, M., Pannu, R., Kanitz, B., & Miville, M. (1998, August). *Kohutian, relational, and racial/cultural theories: Integrative approaches to understanding women.* Symposium presented at the Annual Convention of the American Psychological Association, San Francisco, CA.

Robbins, S. B., & Jolkovski, M. P. (1987). Managing countertransference feelings: An interactional model using awareness of feeling and theoretical framework. *Journal of Counseling Psychology, 34,* 276–282.

Rowe, W., Behrens, J. T., & Leach, M. M. (1995). Racial/ethnic identity and racial consciousness: Looking back and looking forward. In J. Ponterotto, J. M. Casas, & C. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 218–236). Thousand Oaks, CA: Sage.

Sharkin, B., & Gelso, C. J. (1993). The influence of counselor–trainee anger proneness on reactions to an angry client. *Journal of Counseling and Development, 71,* 483–487.

Singer, B. A., & Luborsky, L. L. (1977). Countertransference: The status of clinical versus quantitative research. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: Handbook of research* (pp. 433–451). New York: Pergamon.

Sterba, R. (1934). The fate of the ego in analytic therapy. *International Journal of Psycho-Analysis, 15,* 117–126.

Stolorow, R. (1998). Clarifying the intersubjective perspective: A reply to George Frank. *Psychoanalytic Psychology, 15,* 424–427.

Tang, N. M., & Gardner, J. (1999). Race, culture, and psychotherapy: Transference to minority therapists. *Psychoanalytic Quarterly, 68,* 1–20.

Terrell, F., & Terrell, S. L. (1984). Race of counselor, client sex, cultural mistrust level, and premature termination from counseling among Black clients. *Journal of Counseling Psychology, 31,* 371–375.

Thompson, C. E., & Carter, R. T. (Eds.). (1997). *Racial identity theory: Applications to individual, group, and organizational interventions.* Mahwah, NJ: Erlbaum.

Thompson, C. E., Worthington, R., & Atkinson, D. R. (1994). Counselor content orientation, counselor race, and Black women's cultural mistrust and self-disclosures. *Journal of Counseling Psychology, 41,* 155–161.

Thompson, C. L. (1995). Self-definition by opposition: A consequence of minority status. *Psychoanalytic Psychology, 12,* 533–545.

Van Wagoner, S. L., Gelso, C. J., Hayes, J. A., & Diemer, R. A. (1991). Countertransference and the reputedly excellent psychotherapist. *Psychotherapy: Theory, Research, and Practice, 28,* 411–421.

Worthington, R. L., & Mohr, J. J. (in press). Theorizing heterosexual identity development. *The Counseling Psychologist.*

Political identity and countertransference

Steven Tublin

In the brief essay that follows, I will address a class of transferential and countertransferential experience that receives little attention in the Relational psychoanalytic literature. I am referring to the emotional reactions one experiences when engaging a person whose political identity and values clash with one's own. I will highlight the unique difficulties that such transferential matrices create and will argue that the technical stances adopted to address such conflicts are not likely to resolve these difficulties in treatment.[1]

A clash of political identity – what, for simplicity's sake, I will call "political transference" – is unlike other transference/countertransference matrices. First, to state the obvious, political identity is unlike other identity categories in that it is intrinsically and inescapably adversarial. Being a Democrat or a Republican is therefore not like being a Sagittarius or a Philadelphian. Neither is it fully like "deeper" categories such as male/female or gay/straight, which, though depicting opposite ends of a constructed binary, need not imply hostile opposition. In the American two-party system, being a Democrat or Republican entails, *by definition,* being opposed to the other party – as political figures, as people, and as moral actors in the cultural domain. The two parties exist to engage each other in a battle for power, for a vision of a just society, and for group prominence. And, the more deeply identified one is with one political group, the sharper one's antipathy toward the other party and its members.

Second, I think it is important to recognize that politics, at least for those who hold positions on these matters, is a more fundamental, inner-dwelling component of self-definition than the domain of ideas and opinions which can be altered with relatively little disruption to a person's

sense of self. Political identity is a meaning-giving, morally driven sense of one's place in the world, one that influences not only one's fundamental values and how one sees things, but where and how one lives, who one associates with, how one dresses, what one spends money on, and how one feels about oneself. In a word, one doesn't vote Democrat or Republican so much as one *is* a Democrat or Republican.

And last, the partisan experiences his or her politics with remarkably little friction. However prone to ambivalence a person might be in all other areas of life, politics seems to be felt with utter certainty. One never hears, for instance, "I'm a Democrat, but I hate myself for it." Political identity imposes a clear moral order upon the world, and the partisan knows who the good guys are and what are the proper principles. Political identification thus provides the person with a sense of belonging and an assurance of moral soundness. To say that I am a Democrat means that I am a member of the blue team; they are *my* people. I consider their values to be morally proper, their policies the ones that lead to a more just world, and the members of that group to be better human beings.

The world is thus simplified and the partisan finds he's always on the right side. This is then reinforced in conversation with like-minded group members who share talk of the greed, mendacity, and cognitive limitations of the other guys. Indeed, politics is the only area of contemporary life that I can think of where a kind of conscious splitting is utterly normative. For the partisan, there is little room for doubt, as equivocation carries the whiff of betrayal and risks questions of character. Indeed, venomous disparagement – really *hating* the opposition – connotes a righteous commitment to the just cause.

Committed activists carry this mode of adversarial certainty several degrees further, but the general picture described above could be applied to many citizens who consider themselves politically aligned. This would include most educated, well-informed citizens, a group that would encompass the vast majority of psychoanalytic clinicians and a great many patients. As such, both participants in each analytic dyad often enter the consulting room ready to view the other as either on their team and therefore morally sound and versed in the realities of the world, or on the other team, whose view of the world invites a measure of suspicion or negative judgment. Therefore when analyst's and patient's politics diverge, a morally inflected relational matrix will often emerge that creates an unusual challenge for both parties.

We have become accustomed to working with an array of countertransferential experiences – some intense, some puzzling, some disorganizing – the emergence of which, while unsettling, provides the analyst with information about the patient that is then ideally interwoven into the treatment. Analysts struggle daily with the complex task of remaining open to such experiences while at the same time trying to stand apart from them in order to reflect upon their origins and meaning.

However, political countertransference is often experienced and handled differently. The Relational analyst, working with an assertive, conservative patient, as he talks about his enmity toward Muslims, the corrosive effect of government entitlement programs, or his opposition to same-sex marriage and the rest of the "homosexual agenda," can certainly grow annoyed, infuriated, even disgusted. The registration of a passing hostility toward such a patient is itself unremarkable.

More notable, however, is that there seems to be little inclination with political countertransference for the analyst to thoroughly interrogate his own subjectivity. The analyst in my above example may wonder why the patient is speaking in a way likely to provoke his disdain, or what his own negative feelings might say about the patient's inner world. But, the notion that these feelings and attitudes might be shaped as much by the analyst's history, biases, and particular moral principles – that they are as subjective and relative as any other experience – has not, to my knowledge, been considered.

Rather, the Relational analyst facing the conservative's worldview and values is often thought to be reacting the way any reasonable, decent human being would respond in the face of an objectively noxious stimulus. The analyst might silently wonder, "What experiences led this person to such objectionable positions?" Or, in more explicitly partisan terms, "What led this person to become an awful Republican?" But, what I have not heard asked in this context is: why do Republicans *seem* so awful? And relatedly, how did it happen that Relational clinicians are all Democrats, progressives, or neo-Marxists?

To those on the left, the answer to the first question seems sadly obvious. During the most recent legislative session, the Republican majority in the House of Representatives attempted to cut the budget for food stamps. Left-leaning citizens generally hear this and ask, "What kind of person wants to starve poor children?" The Republican party wants to balance the government's multi-trillion dollar deficit by cutting Social Security and Medicare benefits. Again, the left would ask, "What kind of person wants

to impoverish old people and reduce their medical care?" Republicans (albeit with some increasingly prominent exceptions) seem far more willing than Democrats to employ the ferocious power of the American military throughout the world and seem less concerned with the death and destruction that ensue. To most on the left, this seems incomprehensibly callous. These policies and the people who promote them seem awful because, to those on the left, they simply *are* awful.

The nearly universal leftward tilt among Relational clinicians logically follows. We are all liberals or progressives because we're smart, educated, and compassionate people (or so we like to believe). Many readers, genuinely puzzled, might well ask, "What else would we be?" To the left-leaning partisan, conservatives and Republicans seem to be morally repellent people; in the United States, it simply makes no sense for informed, decent people to be anything other than Democrats or liberals. That conservatives harbor very similar feelings – that *they* are right and good and smart, that the left is populated by the selfish, the ill-informed, the arrogant, and the hypocritical – provokes only fury and confusion on the left.

Of course, some experience this political disjunction far less viscerally. Some Relational analysts live their politics with greater distance. So, while their values may largely align with their more assertive peers, they can put aside at least some political disagreements and retain their stance of general acceptance toward their right-leaning patients.

However, it's hard to know how many accomplish this task, or how they manage it, because no one to my knowledge has written about the issue of political divergence. Rather, the modest literature addressing the issue of politics in the analytic consulting room seems to assume a leftward tilt in the politics of both the analyst and the patient (e.g., Botticelli, 2004; Samuels, 1993, 2004). So seamlessly does this work present the meeting of two left-leaning people that, in reading it, one gets the impression that there are no right-leaning Relational analysts (which is probably not true) and no right-leaning patients (which is obviously not true).

But, absent a thorough sample of politically introspective accounts, it is hard to know how analysts deal with the conflict of values that I am addressing. From anecdotal evidence drawn from conversations with colleagues and students, as well as many intriguingly unguarded posts on online colloquia (where the medium seems to cultivate a fantasy of unanimity and safety which then fosters an unusual candor), my impression is that the approach most commonly employed is to work with the

right-leaning patient as best one can while trying to avoid the tension of mutual political antipathy.

Not surprisingly, those who try this approach report that it's quite difficult. I've heard a number of people describe it in pained and awkward terms, as the silent, if unwelcome, disdain felt toward the patient generates a certain cynicism and a feeling of hopelessness about the treatment. Some analysts describe nevertheless persevering in such circumstances, hoping the treatment will cultivate a greater capacity for empathy in the patient and then somehow a natural evolution to the more "humane" politics of the Democratic party. Beyond the bleak cast imparted to a treatment conducted in this manner, it is worth noting that this approach effectively treats right-leaning politics as intrinsically pathological, a premise arguably true in more extreme cases but one meriting a more nuanced conceptualization in a thorough understanding of political identity.

A less common solution to this problem is to make it a matter of professional policy never to work with Republicans. There is a simple elegance to this stance and certainly all clinicians ought to retain the right not to see a patient they'd prefer not to see. Though, if widely employed, this policy would yield a situation in which Relational psychoanalysis and psychotherapy become treatments offered only by progressives to other progressives.

One occasionally encounters a third approach, in which the analyst, hoping to bring the potentially destructive tension out of the shadows, speaks openly about his political inclinations and the depth of disagreement that exists between him and his patient. One can imagine how this approach might offer the potential to integrate these divergences into the dyadic engagement. But this position too carries considerable risk if presented in a hierarchical or judgmental manner. If, for instance, the analyst proceeds in an evangelical spirit where the analyst endeavors to "teach" the patient the proper politics (e.g., Walls, 2006), one should anticipate suboptimal results. Discussing politics openly but *without* judgment or a wish to influence can be very, very difficult – imagine, for a moment, trying to maintain non-critical curiosity while listening to a patient rail about "libs'" infringement on second amendment rights immediately following the Newtown massacre – but if this issue is to be engaged at all I think this is the only way it can be done.[2]

That is because, in order to navigate this dilemma, the analyst must be prepared to do something that may feel impossible. In order to fully

engage a Republican in treatment, not in the truncated way that treads carefully around his political worldview and identity, the analyst must be prepared to reflect upon his own most basic political assumptions and be ready, not just to learn a thing or two, but to reflect upon his defining values. In a word, the analyst must consider negotiating his non-negotiables (Pizer, 1998), the values he would otherwise never think to question.

There is now, within Relational psychoanalysis, a growing literature employing political theory and constructs to understand patients' suffering, therapeutic change, and technique. This literature is uniformly and unapologetically left in its perspective (cf. Dimen, 2004; Guralnik and Simeon, 2010; Harris, 2012; Layton, 2009; Walls, 2006). Therefore, to get a perspective that looks at political divergence in a non-partisan manner, it is necessary to consult psychological sources outside the domain of psychoanalysis.[3] Here I will be relying primarily on the extensive research of the social psychologist Jonathan Haidt and his colleagues, who have studied the moral underpinnings of various political positions. Haidt (2013) has analyzed several large samples in the United States as well as a variety of demographic groups in India and Brazil. And he has found that moral intuitions rest on five distinct foundations (as he calls them): care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and sanctity/degradation.

All five have at least some resonance with most people, but there are enormous cultural differences, differentiated both by class and geographical region, in the prominence accorded the various foundations in the formation of moral judgment. The end result of this differentiation – and here I am collapsing a library full of social science theory and data – is that around the world and even here in our own country, people rely on widely divergent moral principles to arrive at the values that underlie their politics. Each of these principles is defensible in evolutionary terms and equal in the hold that it has on the psyche of the individual. As a result, we live among people whose morals and politics we find utterly repugnant. And they feel exactly the same way about us.

In particular, Haidt and his colleagues (Graham et al., 2009, 2011; Haidt and Graham, 2007, 2009) found that the politics of the left rely primarily upon two moral foundations: those of care/harm and fairness/cheating. This is why policies diminishing the welfare state are anathema to the left for whom reducing the social safety net beneath the poor and old seems both cruel and unfair and therefore immoral.

But Haidt found that the moral intuitions of people on the right rest on a wider base of principles. As with the left, they do incorporate an ethic of care in their moral intuitions, though they weight it less strongly. And the right also values fairness but, interestingly, they define it differently. Fairness on the left is measured in terms of equality. No one should get too much or too little. The presence in a society of abject poverty juxtaposed with opulent wealth is viewed as evidence of injustice in social policy. However, fairness on the right is measured via what Haidt (2013, p. 169) calls "proportionality." You get what you earn, not more. Inequality is not, in and of itself, immoral.

But the most marked differences come from the right's greater reliance upon the remaining three moral foundations (Haidt and Graham, 2007). Sanctity, for instance, figures prominently in social conservatives' condemnation of many behaviors involving sex or the body in general. One hears the phrases "sanctity of life" and "sanctity of marriage" in debates about social issues. And homosexuality is deplored in religious terms that borrow from the lexicon of purity or decadence. Sanctity as a source of moral judgment barely registers on the left.[4]

There is also, on the right, a far greater reverence for tradition and a wish to protect social institutions such as conventional marriage. Such institutions are viewed with suspicion among the left as instruments for the maintenance of oppressive power relations (thereby violating their ethic of fairness). But among the right they are revered for their ordering of society and the cultural wisdom they are thought to embody.

As it happens, the moral foundations of leftist politics overlap perfectly with the guiding premises of Relational psychoanalysis. Both rest primarily upon empathy, care, and the notion of fairness as defined by equality. Moreover, Relational psychoanalysis retains a subversive, anti-authoritarianism. This theme has been present in psychoanalysis perhaps since Freud and later came front and center in the writing of theorists like Erich Fromm (1941, 1947). That element has been further emphasized in recent Relational work bearing the strong influence of critical and post-structural theory (e.g., Botticelli, 2004; Cushman, 1995, 2009; Dimen, 2004; Harris, 2009, 2012; Layton, 2006, 2009, 2014; Rozmarin, 2010, 2011). This recent work views conformity and adaptation with suspicion and aims to fortify patients in the struggle with what are viewed as oppressive trends (e.g., heteronormativity and rampant consumerism) that exist within broader society. Moreover, like much of progressive politics

and more than other psychoanalytic groups, Relational psychoanalysis, in its moderation of the analyst's role as expert and its emphases on reciprocal influence and mutual inquiry (Hoffman, 1983, 1996), rejects the traditionalism and reverence for authority that so influence the moral intuitions of the right.

As a result, one can expect that, when treating right-leaning patients, the left-leaning Relational analyst will encounter a worldview premised on at least some moral intuitions that seem not moral at all. And because these are deep value judgments, the ensuing countertransference will often lodge in the analyst's psyche in ways that easier differences, easier Othernesses (Haidt et al., 2003), will not. For this reason, I believe therapeutic management of such situations requires much deeper consideration than has been dedicated to date.

In sum, political identity presents a unique countertransferential dilemma in that awareness of it provides no avenue towards its resolution. The visceral hold of political identity, its inherently adversarial orientation, its associated moral certitude, and the evolutionary utility of the group belongingness that accompanies it, create a substantial obstacle to the general acceptance of the other that we regard as a necessary condition for transformative analytic work. In order to perform this work, the analyst must imagine himself having different basic moral principles, in effect not valuing the premises that make him decent and humane in his own eyes. Moreover, in broadening the scope of potential moral intuitions, the analyst risks being shunned within his own community, because psychoanalysis, like everything else, is embedded in world riven by political difference. And in politics, you're either one of us, or you're one of them (Berreby, 2006).

Notes

1 Almost all Relational analysts identify politically left (a situation I address later in this chapter). For that reason, in discussing treatments where the participants' politics diverge, I portray an analyst who identifies somewhere on the left and a patient somewhere right. There are, to be sure, a small number of right-leaning Relational analysts, and I apologize to them as this paper will seem not to address their work experience.

2 Once explicit, it may become apparent to either member of the dyad that the political or moral differences are too stark to overcome and that the extended meeting of *these* two minds is not in the patient's best interest.

3 The field of psychology, like most of the academy today (e.g., Gross, 2013), leans decidedly left (Duarte et al., 2015; Redding, 2001). That skew, while marked, is less pronounced than is observed in Relational psychoanalysis.

4 One does encounter some reliance upon notions of purity and sanctity in the environmental movement and in other narrower concerns regarding food supply.

References

Berreby, D. (2006). *Us and Them: Understanding your Tribal Mind*. New York: Little, Brown.

Botticelli, S. (2004). The politics of relational psychoanalysis. *Psychoanalytic Dialogues*, 14: 635–651.

Cushman, P. (1995). *Constructing the Self, Constructing America: A Cultural History of Psychotherapy*. Reading, MA: Addison-Wesley.

Cushman, P. (2009). Empathy – what one hand giveth, the other taketh away: commentary on paper by Lynn Layton. *Psychoanalytic Dialogues*, 19: 121–137.

Dimen, M. (2004). The return of the dissociated: discussion. *Psychoanalytic Dialogues*, 14: 859–865.

Duarte, J., Crawford, J., Stern, C., Haidt, J., Jussim, L., and Tetlock, P. (2015). Political diversity will improve social psychological science. *Behavioral and Brain Sciences*, 38: 1–13.

Fromm, E. (1941). *Escape From Freedom*. New York: Avon Books.

Fromm, E. (1947). *Man for Himself: An Inquiry into the Psychology of Ethics*. New York: Holt, Rinehart, and Winston.

Graham, J., Haidt, J., and Nosek, B. (2009). Liberals and conservatives rely on different sets of moral values. *Journal of Personality and Social Psychology*, 96: 1029–1046.

Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., and Ditto, P. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101: 366–385.

Gross, N. (2013). *Why Are Professors Liberal and Why Do Conservatives Care?* Cambridge, MA: Harvard University Press.

Guralnik, O., and Simeon, D. (2010). Depersonalization: standing in the spaces between recognition and interpellation. *Psychoanalytic Dialogues*, 20: 400–416.

Haidt, J. (2013). *The Righteous Mind: Why Good People Are Divided by Politics and Religion.* New York: Pantheon Books.

Haidt, J., and Graham, J. (2007). When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20: 98–116.

Haidt, J., and Graham, J. (2009). Planet of the Durkheimians, where community, authority, and sacredness are foundations of morality. In J. Jost, A. C. Kay, and H. Thorisdottir (Eds.), *Social and Psychological Bases of Ideology and System Justification*. New York: Oxford University Press, pp. 371–401.

Haidt, J., Rosenberg, E., and Hom, H. (2003). Differentiating diversities: moral diversity is not like other kinds. *Journal of Applied Social Psychology*, 33: 1–36.

Harris, A. (2009). You must remember this. *Psychoanalytic Dialogues*, 19: 2–21.

Harris, A. (2012). The house of difference, or white silence. *Studies in Gender and Sexuality*, 13: 197–216.

Hoffman, I. (1983). The patient as interpreter of the analyst's experience. *Contemporary Psychoanalysis*, 19: 389–422.

Hoffman, I. (1996). The intimate and ironic authority of the psychoanalyst's presence. *Psychoanalytic Quarterly*, 65: 102–136.

Layton, L. (2006). Racial identities, racial enactments, and normative unconscious processes. *Psychoanalytic Quarterly*, 75: 237–269.

Layton, L. (2009). Who's responsible? Our mutual implication in each other's suffering. *Psychoanalytic Dialogues*, 19: 105–120.

Layton, L. (2014). Grandiosity, neoliberalism, and neoconservatism. *Psychoanalytic Inquiry*, 34: 463–474.

Pizer, S. (1998). *Building Bridges: The Negotiation of Paradox in Psychoanalysis*. Hillsdale, NJ: The Analytic Press.

Redding, R. (2001). Sociopolitical diversity in psychology. *American Psychologist*, 56(3): 205–215.

Rozmarin, E. (2010). Better identity politics. *Psychoanalytic Dialogues*, 20: 181–190.

Rozmarin, E. (2011). To be is to betray: on the place of collective history and freedom in psychoanalysis. *Psychoanalytic Dialogues*, 21: 320–345.

Samuels, A. (Ed.) (1993). *The Political Psyche*. London: Routledge.

Samuels, A. (2004). Politics on the couch? Psychotherapy and society – some possibilities and some limitations. *Psychoanalytic Dialogues*, 14: 817–834.

Walls, G. (2006). The normative unconscious and the political contexts of change in psychotherapy. In L. Layton, N. C. Hollander, and S. Gutwill (Eds.), *Psychoanalysis, Class and Politics*. New York, NY: Routledge, pp. 118–128.

ATTITUDE TRANSFERENCE

MARK D. ALICKE, FRANK M. LOSCHIAVO, AND JUSTIN T. BUCKINGHAM
Ohio University

Attitude transference involves transferring an attitude expressed by one person onto other, neutral targets. This attitude transference phenomenon was investigated in five studies using variations of the attitude-attribution paradigm. Participants heard a target express either favorable or unfavorable attitudes on a topic and were subsequently asked to guess the attitudes of neutral targets in the context of a different experiment. Path analyses showed that exposure to favorable or unfavorable attitudes directly influenced attitude attributions, and that neither heightened consensus estimates nor priming of personal attitudes could completely account for the results. The implications of attitude transference for persuasion and impression formation research were discussed.

The idea that cognitions evoked by one source can be transferred onto another underlies various psychological theories. Perhaps the most prevalent cognitive transference theory is projection. Freud introduced the term "projection" into the psychology vernacular to represent a defense mechanism that unconsciously deflects threatening thoughts, motives, or emotions from oneself to others (1956). Neo-Freudians such as Horney (1939) broadened the projection concept to include any tendency to attribute one's mental contents to others, including both positive and negative contents. In an influential literature review, Holmes (1968) concluded that there was ample empirical evidence for projection, but little support for the Freudian assumption that projection occurs without awareness.

A second type of cognitive transference involves "social projection" (Krueger, 1998). Social projection is the attribution of one's attitudes, values, or behavioral tendencies onto social groups, particularly peer groups. Specifically, social projection entails a correlation between one's attitudes or propensities and estimates of others' attitudes (Katz &

Allport, 1931). Interest in social projection proliferated with the publication of Ross, Greene, and House's (1977) paper on the "false consensus effect." The basic false consensus paradigm asks participants to make dichotomous choices between attitude options or behavioral preferences (e.g., liking vs. disliking lima beans; willingness vs. unwillingness to donate money to charity), and then asks them to estimate the percentage of their peers who would endorse each item. The typical results of false consensus studies show that people overestimate consensus for their own views. For example, lima bean afficionados overestimate the percentage of their peers who like lima beans relative to lima bean detractors.

Skowronski, Carlston, Mae, and Crawford (1998) have recently identified a third type of cognitive transference, which they call trait transference. In their research, participants who read sentences in which targets described acquaintances with trait-implying sentences subsequently associated the targets with those traits. In other words, participants transferred the traits from the people who were described to the targets who provided the descriptions. Other studies have replicated and extended these findings (D'Agostino & Hawk, 1998).

Research on category-activation, or priming, can also be construed as a form of cognitive transference. In a prototypical study by Higgins, Rholes, and Jones (1977), participants who were primed with a trait such as "adventurous" subsequently formed more favorable impressions of a person who sky-dived and drove in demolition derbies than participants primed with the trait reckless. In this type of priming study, participants use an activated construct to resolve ambiguous behavior. In the terminology of cognitive transference, the trait to which participants are initially exposed is transferred to the target whose behavior it "disambiguates."

The four main components of cognitive transference are delineated in Table 1. These components include the source that generates the item to be transferred, the content of that item, a target onto whom the item is transferred, and the content of the transferred item. In classical projection, for instance, the exposure source is the self; the content of the item is a thought, motive, belief, or emotion; and the transference target is another person. As Holmes (1968) noted, the transferred contents in classical projection are either identical with the exposure contents (attributive projection) or complementary to those contents (complementary projection). Projecting identical contents is exemplified by a frightened person seeing others as frightened, whereas complementary projection entails a frightened person seeing others as frightening (Murray, 1933).

In social projection, the generative source is again the self, the contents to be transferred are one's attitudes or behavior preferences, and these attitudes or preferences are transferred onto a reference group. In trait

TABLE 1. Varieties of Cognitive Transference

	Generative Source	Content of Item to be Transferred	Transference Target	Content of Transferred Item
Classical Projection	Self	Motives or Beliefs	Person A	Same motive or belief (attributive projection) or its complement (complementary projection)
Social Projection	Self	Attitudes or Behavior Preferences	Peer Group	Same attitude or behavior preference
Trait Transference	Person A	Trait-Implying Sentence	Person A	Trait implied by sentence
Priming	External Source	Trait	Person A	Trait
Attitude Transference	Person A	Attitude	Person B	Same attitude

transference, a trait implied in a behavior description is transferred back to the source who provided the description. Finally, in priming or category activation, people are exposed to a trait by an external source (e.g., by a computer-generated presentation), and this trait is then transferred onto another person after being used as a framework for interpreting the person's behavior.

In this article, we focus on a novel form of cognitive transference that we call attitude transference. Attitude transference occurs when an attitude expressed by one person is transferred onto another. For example, after hearing Person A express an unfavorable view of government sponsored welfare, the person who heard the communication attributes that attitude to Person B. This type of transference differs from classical and social projection in that the generative source is someone other than oneself. Furthermore, attitude transference differs from trait transference in that the content is transferred from one source to another rather than reflexively attributed to the original source. Finally, whereas in most priming research the transferable content is not initially associated with a person, in attitude transference, participants transfer the content from one personal source to another. In priming studies, however, an activated category or schema serves as an interpretational framework for resolving ambiguous information. Activating a minority group stereotype, for example, may lead observers to interpret ambiguous behavior in a way that confirms the stereotype. In contrast to priming, attitude transference involves a direct attribution of a content (i.e., an attitude) to a neutral target about whom no specific information is available. Thus, attitude transference raises the novel possibility that simply hearing an

attitude expressed by one person will increase the likelihood of that position being ascribed to another.

A secondary purpose of this article is to suggest that cognitive transference may provide a partial explanation for correspondence bias (Jones & Harris, 1967). Participants in correspondence bias studies are typically exposed to a target who argues the favorable or unfavorable side of an attitude issue (e.g., pro- or anti-marijuana legalization). Some participants are told that the targets freely expressed their attitudes, whereas others learn they were randomly assigned to express these views. Participants then estimate the target's true attitude. Studies using the attitude attribution paradigm confirm correspondent inference theory's prediction that participants make more extreme attitude attributions when they believe targets expressed their views freely rather than under constraint. However, the fact that a significant difference in attitude attributions remains in the assignment condition shows that people do not correct sufficiently for the constraint. In other words, although participants realize the attitude positions were randomly assigned, they nevertheless attribute attitudes consistent with the favorable or unfavorable views expressed (e.g., Jones & Harris, 1967; Miller, Ashton, & Mishal, 1990; Snyder & Jones, 1974).[1]

In the attitude attribution paradigm, the target who expresses an attitude is the generative source of the attitude. According to the cognitive transference scheme described in Table 1, this attitude may be used as a basis for estimating the attitude of the target who generated it or the attitude of a different target. The target's expression in the attitude attribution paradigm is invalidated as an indicator of the target's true beliefs by the fact that it was randomly assigned by the experimenter. Similarly, in trait transference, the target's expression is an invalid indicator of the target's traits because the expression pertains to another person. In both paradigms, however, participants fail to discount the invalidated expression and attribute it to the generative source.

We believe that with all other things being equal, people tend to transfer contents that have been recently generated to designated targets—regardless of whether the generated contents represent true beliefs or are randomly assigned. Furthermore, this transference is expected to occur

1. It is frequently said that people should discount personal characteristics in the presence of strong situational constraints. Strictly speaking, this is untrue. The fact that a person loses her temper while in the throes of a migraine headache does not mean that she is *not* an angry person, it simply ambiguates the diagnosticity of the behavioral evidence. The basic assumption of the covariation principle underlying attribution theory is that to assess her personality we would have to analyze her behavior in the presence and absence of migraine headaches.

regardless of whether the generative source is oneself or another person, and regardless of whether the designated target is the original source or a different person. In the attitude attribution paradigm, the generative source is the person who generated the attitude statement and the designated target is the same person. However, if correspondence bias can be at least partly explained in terms of people transferring recently generated attitude statements to designated targets, then a similar effect should emerge when participants ascribe attitudes to neutral sources, that is, to people who never expressed any attitudes. Participants who are exposed to a source who expresses a pro-marijuana attitude, for example, should ascribe more favorable views of marijuana to neutral targets than participants who are exposed to a source who expresses an anti-marijuana attitude.

Attitude transference was studied in five separate experiments. Participants in Study 1 read 15 favorable or 15 unfavorable attitude statements written by a student on the topic of standardizing college courses. Some participants believed the student's views were freely expressed whereas others believed they were assigned. Participants were then asked to estimate the true attitudes of six neutral targets whose images appeared on their computer screens. Studies 2 and 3 were more direct analogues of previous attitude attribution studies. Participants watched videotaped segments in which a target argued either in favor of or against homosexuals serving in the military, and then estimated the attitudes of five people whose faces they saw on a TV screen. Study 4 incorporated response-time measures to assess the role of attitude accessibility in attitude transference. Finally, Study 5 was designed to show that the attitude transference effect perseveres, while accounting for the possibility that participants' own attitudes are altered by exposure to the targets' statements and then transferred to the designated targets.

STUDY 1

METHOD

Participants. Participants in Study 1 were 111 (48 male and 63 female) undergraduate students whose participation partially fulfilled a course requirement.

Procedure. Upon entering the laboratory, participants were seated at computers and presented with the following instructions on their screens:

> In this experiment, you will evaluate some statements that have been developed for a public service announcement. All the information you will need

will be provided by a computer program. It is important that you take your time and read all the information and instructions carefully. The entire program will take only about 10 minutes so relax and proceed at your own pace.

The experimenter showed participants how to navigate through the program and then asked them to begin. The program began by elaborating the cover story regarding the public service announcement. Participants read the following information:

As you may already know, Ohio University's Scripps School of Journalism is one of the finest and best known journalism schools in the country. Because of the school's noteworthiness, faculty and graduate students are often offered freelance writing opportunities. The National Education Foundation has recently hired a team composed of faculty and graduate students from Scripps to write a public service announcement regarding the standardization of college courses. To ensure that their announcement conveys the message they intend, the Scripps team has hired some researchers in the Psychology Department to test its content.

During this experiment, your task will be to read several statements that have been generated for the public service announcement. Furthermore, after reading the statements, you will be asked to respond to a series of questions regarding the standardization issue. So that you better understand what course standardization entails, you will now have the opportunity to read some information about it.

Many educators have suggested that it would be a good idea to standardize certain college courses, especially those required for graduate and professional schools. One example is the sequence of biology courses required for pre-medical students. It has been argued that the courses in this sequence should cover standardized material and that test items should all be selected from a central test bank. If this were done, students at Ohio University would take the same biology classes, and essentially the same biology tests, as students at other universities. Based on the data we have collected so far, about half of all college students favor the idea of course standardization and about half oppose the idea.

The free choice versus constraint manipulation was introduced at this point. Participants in the free choice condition read the following information:

The National Education Foundation wants the Scripps team to develop an announcement that covers both favorable and unfavorable aspects of course standardization. The goal of the announcement is to inform the public of the relevant issues. The statements you will evaluate today have been written by one member of the Scripps team. Like every member of the team,

this person generated several statements that could be used for the public service announcement. Because both favorable and unfavorable statements were needed, this writer was free to generate any type of statement. Therefore, the writer wrote statements that expressed his or her true attitudes.

Participants in the constraint conditions read a slightly modified version:

The National Education Foundation wants the Scripps team to develop an announcement that strongly favors (opposes) course standardization. The goal of the announcement is to convince the public that standardization is (is not) desirable. Because the National Education Foundation strongly favors (opposes) standardization, only statements that favor (oppose) standardization were needed. Therefore, the writer was hired to generate statements that favor (oppose) standardization.

Attitude Statements. The favorable or unfavorable statements students read were selected from a pool of attitude statements on course standardization generated by 107 students enrolled in a standardized statistics course. These students were briefed on the issue and asked to generate a range of positive and negative statements about standardization. In exchange for course credit, a separate group of 45 introductory psychology students rated each statement on an 11-point scale ranging from -5 (extremely unfavorable) to +5 (extremely favorable). Average ratings were computed for each statement. We selected the 15 most favorably rated statements and the 15 most unfavorably rated statements.

Dependent Measures. After reading all the statements, participants predicted the attitudes of six neutral targets whose digitized images were presented on their computer monitor. In addition, participants estimated the percentage of their peers who they believed supported or opposed standardizing courses. Participants were told to be sure that their percentage estimates totaled 100%. All ratings (except for consensus estimates) were made on 9-point scales ranging from *totally opposes the issue* (1) to *totally supports the issue* (9).

RESULTS AND DISCUSSION

No effects due to subjects' gender were obtained in this or in subsequent studies, so this factor was eliminated from the analyses.

Manipulation Checks. The first manipulation check assessed whether participants recognized that the attitude propositions they read were freely expressed or assigned. This difference was significant, $F(1, 109) =$

TABLE 2. Mean Attributions and Standard Deviations for Neutral Targets' Beliefs
About Course Standardization (Study 1)

	Free Choice		Assignment	
	Pro	Anti	Pro	Anti
Attitude Attributions				
M	5.87	5.03	5.81	4.72
SD	0.84	1.06	0.84	0.76
n	29	27	30	25
Consensus Estimates				
M	52.62	46.96	53.83	44.44
SD	10.73	10.57	11.95	8.08
n	29	27	30	25

42.48, $p < .0001$, showing that participants in the free choice condition
perceived the targets to be freer to express their true views ($M = 7.70$, SD
$= 1.65$) than did participants in the assignment condition ($M = 5.04$, $SD =$
2.56). The second manipulation check assessed whether participants
were aware of the favorable versus unfavorable nature of the proposi-
tions. This manipulation check was also successful, $F(1, 109) = 448.44$, $p <$
.0001, showing that participants exposed to favorable propositions
viewed those propositions as more favorable ($M = 8.00$, $SD = 1.07$) than
did participants exposed to unfavorable propositions ($M = 2.08$, $SD =$
1.82).

Attitude Attributions. Attitude attributions to the six neutral targets
were combined to form a composite index.[2] The means and standard de-
viations are shown in Table 2. The main analysis involved the attitudes
attributed by participants exposed to attitude propositions favorable
and unfavorable toward standardized courses. This analysis was first
conducted with attitude direction (favorable or unfavorable) and choice
(free choice or assignment) as the two independent variables. No main
effects or interactions were obtained for the choice manipulation. Over-
all, the effect of attitude direction was significant, $F(1, 107) = 33.00$, $p <$
.0001, indicating that people exposed to favorable positions ascribed

---

2. Cronbach's alpha was in the low to moderate range in these studies (Study 1: $\alpha = .46$;
Study 2: $\alpha = .45$; Study 3: $\alpha = .14$; Study 4: $\alpha = .37$; Study 5: $\alpha = 43$). The lowest value, in Study
3, was obtained after a 1-week delay. These data suggest that there was considerable varia-
tion in attributions among the six neutral targets, and that factors (such as perceptions of
nonverbal cues) other than accessibility influenced participants' judgments. We do not
claim, of course, that accessibility is the only factor which influences neutral target judg-
ments in this situation. However, the fact that the accessibility effect is consistently ob-
tained, even with considerable variability among the six targets, suggests that the effect is
relatively robust.

more favorable attitudes (M = 5.84, SD = 0.83) than those exposed to un-favorable positions (M = 4.88, SD = 0.93). Although this finding did not interact with the choice manipulation, we analyzed the effect within the assignment condition because the correspondence bias phenomenon pertains specifically to differences in ascribed attitudes in this condition. As predicted, the usual correspondence bias effect was obtained, $F(1, 107) = 20.87$, $p < .0001$, although the targets for whom the attitudes were ascribed were not the ones who originally expressed the attitudes.

Consensus Estimates. Based on research showing that a single item of case information can disproportionately influence social projections (Alicke & Largo, 1995; Sherman, Presson, & Chassin, 1984), it seems likely that reading attitude propositions on one side of an issue would lead participants to overestimate consensus for that position. The tendency to overestimate consensus for the attitude expressed suggests that hearing an attitude position expressed may lead participants to overestimate consensus for that attitude, and attitude attributions to neutral targets may be based solely on heightened consensus. If so, heightened consensus would provide a sufficient explanation for attitude transference.

As expected, we obtained a significant effect of attitude position on consensus information, $F(1, 107) = 14.14$, $p < .0003$. The consensus effect showed that participants who read favorable attitude propositions esti-mated that a greater percentage of their peers held favorable attitudes toward standardization (M = 53.24, SD = 11.28) than those who read un-favorable propositions (M = 45.75, SD = 9.45). A path analysis was there-fore conducted to assess the direct and indirect effects of attitude position on attitude attributions. (Because we predicted positive rela-tionships in all instances, one-tailed significance tests were used for all path analyses in this and in the following studies.) The results of this analysis are provided in Figure 1. As Figure 1 shows, exposure to favor-able or unfavorable attitude positions influenced attitude attributions both directly as well as indirectly via their influence on consensus esti-mates. The significant direct effect of attitude position suggests that heightened consensus cannot completely explain attitude transference.

In contrast to most correspondence bias research, attitude polarization (i.e., the difference between attitudes attributed to targets who ex-pressed favorable vs. unfavorable views) was approximately equal in the free choice and assignment conditions. The most likely explanation for this difference stems from the fact that attributions in this study were made for neutral targets who did not express any attitudes, and who were under no constraints rather than for the original target. The fact that the attitude transference effect was approximately equal in the free choice and constraint conditions means either that the effects of attitude exposure were sufficiently strong to override the effects of hearing a con-
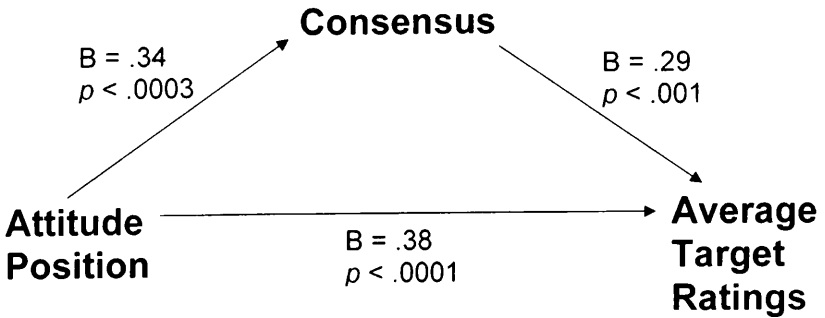
Consensus

B = .34
p < .0003

B = .29
p < .001

**Attitude
Position**

B = .38
p < .0001

**Average
Target
Ratings**

FIGURE 1. Mediation analysis (Study 1).

Note. Standardized beta weights are reported.

strained presentation or that the constraint was relatively weak. We address this issue in Study 5 by employing a stronger constraint manipulation.

Study 1, therefore, showed that exposing participants to opposite sides of an attitude issue led them to attribute this attitude to neutral targets. Furthermore, this effect could not be completely explained in terms of perceived consensus. To control the stimulus presentation as precisely as possible, the methodology of Study 1 diverged from the usual correspondence bias paradigm. In most correspondence bias studies, a single target writes an essay or gives a speech espousing a favorable or unfavorable attitude position. We, on the other hand, provided 15 discrete favorable or unfavorable attitude propositions on the computer. Our methodology could conceivably have divorced the message from the source to a greater extent than in most correspondence bias studies, and therefore have increased the tendency to transfer correspondent attitudes to neutral sources. The main purpose of Study 2 was to investigate attitude transference with the traditional attitude attribution paradigm.

STUDY 2

Participants in Study 2 were exposed to three students who expressed their views on a variety of controversial topics, and then to a fourth student who expressed his or her views on the critical attitude issue—whether homosexuals should be permitted to serve in the military. Participants were told either that the students' views were freely ex-

pressed or that they were assigned to express favorable or unfavorable attitudes. After hearing the attitude expressed, participants were brought to another building for an ostensibly unrelated experiment involving impression formation based on nonverbal cues. Participants were asked a variety of questions about five people whose faces they saw on a TV screen. Embedded within these questions was the critical item asking them to guess the targets' attitudes toward homosexuals serving in the military. We predicted that participants who originally heard favorable views expressed toward homosexuals in the military would ascribe more favorable views to neutral targets than would participants who originally heard unfavorable views expressed.

METHOD

Participants and Design. Eighty-nine undergraduate students (47 male and 42 female) participated in two ostensibly unrelated experiments in partial fulfillment of a course requirement.

Procedure and Manipulations. Phase 1. Participants were led to believe they would be serving in two separate experiments. They were told the first experiment would last about 15 minutes and that upon completion they would be escorted to another building for another, unrelated study.

The cover story conveyed that participants would be observing videotaped speeches delivered by journalism students as part of a class project. Participants in the assignment condition were told that the professor assigned a speech requiring students to argue in favor of or against a designated issue. The instructions emphasized that speakers were not allowed to choose which side of the issue they would support. Participants in the free choice condition were told that students were free to choose whichever side of the attitude issue conveyed their true feelings.

The experimenter then played a videotape of four students presenting speeches on the exploitation of women, the legal drinking age, abortion—and the critical issue—whether homosexuals should be permitted to serve in the military. The attitudes conveyed by the first three speakers were ambiguous in that the speakers provided arguments on both sides of each issue. The purpose of providing ambiguous presentations was to avoid creating a valence effect such that participants might transfer positive or negative positions instead of the specific attitude issue. Participants were then randomly assigned to hear a speaker unambiguously favoring or opposing homosexuals in the military. In half the cases, this speaker was female, and in the other half, male. The issue regarding homosexuals serving in the military was always presented last to ensure that propositions either favoring or opposing that topic were the most accessible.

TABLE 3. Mean Attributions and Standard Deviations for Neutral Targets' Beliefs About Homosexuals Serving in the Military (Study 2)

	Free Choice			Assignment	
	Pro	Anti		Pro	Anti
Attitude Attribution					
M	5.32	4.50		5.40	4.53
SD	0.85	1.03		0.94	1.01
n	23	23		21	22
Consensus Estimates					
M	57.61	42.83		52.86	39.00
SD	20.06	16.91		21.13	18.18
n	23	23		21	22

In keeping with the cover story, participants responded to questions evaluating several dimensions of each speech and speaker. Embedded within this evaluation form, participants reported their estimates of the percentage of college students who supported each side of the issue.

Phase Two. The experimenter then brought participants to an ostensibly unrelated experiment in a different building. After obtaining additional informed consent, a second experimenter explained that the next study was concerned with impression formation and nonverbal communication. Participants viewed videotaped images of five different target people (3 male and 2 female) for 10 seconds each. The order of presenting the targets was randomized. After viewing each image, the tape was stopped and participants rated the target's attitude on several current issues. Embedded within several questions, participants estimated the extent to which each target person supported or opposed homosexuals serving in the military. After completing this task, participants were probed for suspicion and debriefed.

RESULTS AND DISCUSSION

Attitude Attributions. Attitude attributions to the five targets were combined to form a composite index. The means and standard deviations for average attitude ratings and consensus estimates are shown in Table 3. As in Study 1, no main or interaction effects were obtained due to the choice manipulation ($Fs < 1$). As predicted, however, students exposed to the pro-homosexual speech rated targets significantly more in favor of homosexuals serving in the military ($M = 5.36, SD = .89$) than did students exposed to the anti-homosexual speech ($M = 4.51, SD = 1.01$), $F(1, 85) = 17.54, p < .0001$. As in Study 1, we examined this effect within
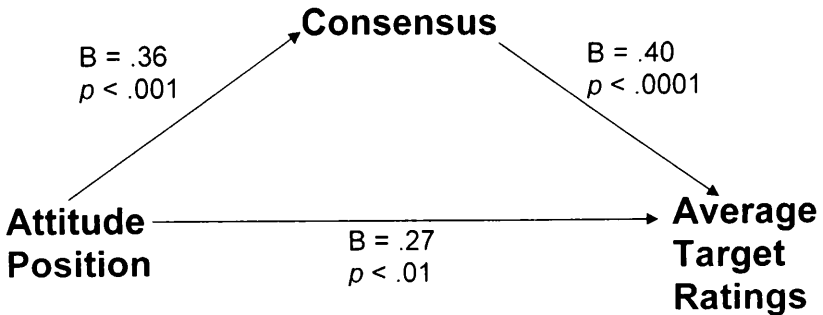
FIGURE 2. Mediation analysis (Study 2).

Note. Standardized beta weights are reported.

the assignment condition for consistency with prior research. This effect was also significant, $F(1, 85) = 8.95, p < .004$, showing again that the basic correspondence bias effect extends to targets who did not originally express any attitudes.

Consensus Estimates and Path Analysis. As in the first study, participants exposed to the pro-homosexual speech estimated that more students favored this position than did participants exposed to the anti-homosexual speech, $F(1, 85) = 12.49, p < .001$. A path analysis was again conducted to assess the direct and indirect influences of attitude position on attributed attitudes. The results are shown in Figure 2. As in Study 1, exposing participants to favorable or unfavorable attitudes directly influenced attitude attributions, and also affected attributions indirectly by influencing perceived consensus. These findings show again that perceived consensus cannot completely explain the attitude transference effect.

STUDY 3

The two main purposes of Study 3 were to replicate the second study's findings using the typical correspondence bias paradigm, and to assess whether attitude transference effects persevere after a 1-week delay. Showing that transference effects persevere for at least a week after the original manipulation would demonstrate the robustness of people's tendencies to transfer attitude positions generated by one target to neutral targets.

METHOD

Participants. Thirty-two undergraduate students (11 male and 21 female) participated in two ostensibly unrelated experiments in partial fulfillment of a course requirement.

Procedure. The procedure in Study 3 paralleled that of Study 2, with one major exception. Students in this experiment were scheduled to participate in two ostensibly unrelated experiments to be held approximately one week apart. Experimenters telephoned students and offered them the opportunity to participate in several studies; only students interested in scheduling two appointments were contacted.

As in Study 2, the first part of the experiment required participants to evaluate four speeches. Participants were again told that a professor assigned a speech requiring students to argue in favor of or against a designated issue, and that speakers were not allowed to choose which side of the issue they would support. No free choice condition was included in this study.

Approximately one week later (Mean = 5.7 days, Mode = 7 days), participants took part in what they believed was an unrelated study. Each participant viewed videotaped images of five different target people and rated each target's attitude on several dimensions, including the extent to which each supported or opposed homosexuals serving in the military. In addition, participants provided consensus estimates of college students' attitudes on this dimension.

RESULTS AND DISCUSSION

Attitude Attributions. Table 4 displays the means and standard deviations for participants' attitude attributions and consensus estimates. As in the previous studies, we obtained a significant effect of attitude position on neutral target attributions,[3] $F(1, 30) = 4.27, p < .05$. Thus, even after a one-week delay, students exposed to the pro-homosexual speech rated neutral targets as significantly more in favor of homosexuals serv-

3. Study 3 was the only one in which we recorded participants' responses to the filler items in addition to the critical issue. The two filler questions were "To what extent do you think this student is in favor of or against federal health care reform?" and "To what extent do you think this student is in favor of or against government regulations concerning the amount of violence on TV?" No evidence of attitude projection was obtained on either of these questions ($F < 1$ in each instance), thereby showing that attitude projection occurred only on the critical issue (i.e., gays in the military). This result is important because it rules out the possibility that people were simply projecting positive or negative views instead of the specific contents of the original target's message.

TABLE 4. Mean Attitude Attributions, Standard Deviations, and Consensus Estimates for Original and Neutral Targets' Beliefs About Homosexuals Serving in the Military (Study 3)

	Attitude Expressed	
	Pro	Anti
Attitude Attribution		
Neutral Targets (after 1-week delay)		
M	5.63	4.99
SD	0.95	0.79
n	16	16
Consensus Estimates (directly after priming procedure)		
M	55.00	33.06
SD	19.66	16.25
n	16	16
Consensus Estimates (after 1-week delay)		
M	51.88	40.63
SD	17.40	19.82
n	16	16

ing in the military than did students exposed to the anti-homosexual speech.

Consensus Estimates and Path Analysis. Although consensus estimates did not differ based on pretest measures ($F < 1$), posttest consensus estimates (obtained directly after attitude exposure) differed significantly, $F(1, 30) = 6.25, p < .02$. As Table 4 shows, students exposed to the pro-homosexual speech estimated that more students favored homosexuals serving in the military than did students exposed to the anti-homosexual speech. This effect was maintained even after one week, $F(1, 30) = 2.91, p < .05$. These data were again elaborated with a path analysis as shown in Figure 3. This analysis used attributions and consensus estimates measured after a one-week delay.[4] In contrast to the first two studies, the attitude position to which participants were exposed exerted only a direct effect on attitude attributions; in this case, no mediated path with consensus information was obtained. In this study, therefore, the dissipation of the consensus effect over time removes consensus as an explanation of attitude transference.

4. We also conducted this analysis using consensus estimates from Time 1 (measured immediately after exposure to the attitude expression). As with the delayed measure, no significant path was obtained from consensus to average target ratings.
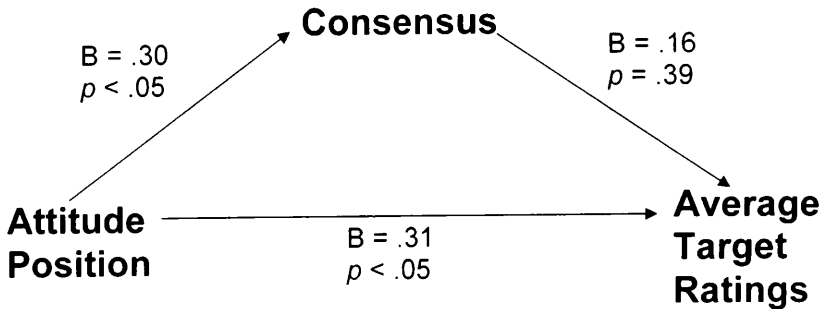
FIGURE 3. Mediation analysis (Study 3).
Note. Standardized beta weights are reported.

STUDY 4

The first three studies demonstrated that hearing favorable or unfavorable views led participants to transfer correspondent attitudes to neutral targets. The first two studies also suggested that perceived consensus is one factor that contributes to the attitude transference phenomenon. However, perceived consensus is not a sufficient explanation in that the attitude transference effect occurs when the effects of perceived consensus are statistically removed.

Study 4 assessed another possible contributor to attitude transference, namely, cognitive accessibility. According to this possibility, hearing predominantly favorable or unfavorable views heightens the accessibility of favorable and unfavorable attitudes respectively and increases the likelihood of their being transferred to neutral targets. Accessibility was measured by exposing participants to favorable or unfavorable attitude statements and asking them to respond as quickly as possible to a new set of favorable and unfavorable statements. Direct evidence of heightened accessibility would result in an interaction showing that participants exposed to favorable attitude statements respond more quickly to the new set of favorable relative to unfavorable statements, whereas those exposed to unfavorable attitude statements respond more quickly to the new set of unfavorable relative to favorable statements.

Another purpose of Study 4 was to compare attributions made to the original target with those made to neutral targets. Because we were primarily interested in demonstrating attitude transference effects, we did not ask participants in the first three studies to attribute attitudes to the

original target. Original target attributions, however, provide a useful comparison for evaluating the neutral target effect. Despite random assignment, participants may perceive a unit relationship between the person who expresses the attitude and the position the attitude expresses (Heider, 1958). As Gilbert and Malone (1995) have argued, people automatically assume the statements they hear others express represent their true opinions, an assumption that applies in our study to the original target, but not to neutral targets who do not express any views. Thus, the original target is likely to be viewed as the "owner" of his or her attitude expressions and, for this reason, attributions to the original target are likely to be more extreme than for the neutral targets. Accordingly, we expected that correspondence bias would be stronger for the original target than for neutral targets. More specifically, we predicted an interaction such that the difference in attributed favorable and unfavorable attitudes would be larger for the original target than for the neutral targets.

METHOD

Participants. One hundred twenty-two undergraduate students participated in partial fulfillment of a course requirement.

Procedure. The main procedure of the fourth study was the same as the first (i.e., statements regarding course standardization were displayed on a computer screen), except that participants estimated the true attitude of the original target as well as the neutral targets and the free choice condition was excluded. Study 4 also included measures designed to assess the accessibility of propositions associated with favorable and unfavorable attitude statements. After reading the statements and rating the original and neutral targets, participants were presented with 20 additional statements, half favoring and half opposing course standardization. Positive and negative statements were pretested to be approximately equal in length. Participants were instructed to decide as quickly and accurately as possible whether each statement favored or opposed course standardization. Responses were registered by pressing one of two keys representing favorable or unfavorable responses. Participants placed an index finger on each key and kept their fingers in that position throughout the experiment. We recorded the time it took them to respond to a statement from its onset, and also whether they correctly identified if the statement supported or opposed the standardization issue. If a participant made a correct judgment, the next trial began. If the participant made an incorrect judgment, a beep indicated that an error had been made before the next trial began. Following this response time procedure, participants again estimated the attitudes of six neutral tar-

TABLE 5. Mean Attributions and Standard Deviations for Original and Neutral Targets' Beliefs About Course Standardization (Study 4)

	Attitude Expressed	
	Pro	Anti
Attitude Attribution		
Original Target		
M	7.68	2.33
SD	1.30	1.89
n	62	60
Neutral Targets		
M	5.89	5.16
SD	0.73	1.02
n	62	60
Consensus Estimates		
M	54.63	44.60
SD	12.71	10.80
n	62	60

gets whose pictures were presented on the computer screen as well as the attitudes of the original target.

RESULTS AND DISCUSSION

Attitude Attributions. Table 5 displays means and standard deviations for attitude attributions made to both the original and neutral targets. Original and neutral targets were included as a within-subjects factor in an analysis of variance, along with the between-subjects factor of attitude direction (favorable or unfavorable). The basic correspondence bias effect was obtained for both the original and the neutral targets. Regarding the original target, results showed that participants exposed to favorable statements ascribed more favorable attitudes to the writer (M = 7.68, SD = 1.30) than did those exposed to unfavorable statements (M = 2.33, SD = 1.89), $F(1, 120) = 331.87, p < .0001$. This finding is similar to the one usually obtained in attitude attribution studies and shows participants attributed correspondent attitudes to the original writer despite random assignment.

The same analysis was significant with neutral targets, $F(1, 120) = 21.01, p < .0001$, indicating that participants exposed to favorable positions ascribed more favorable attitudes to the neutral targets (M = 5.89, SD = 0.73) than did participants exposed to unfavorable positions (M = 5.16, SD = 1.02).

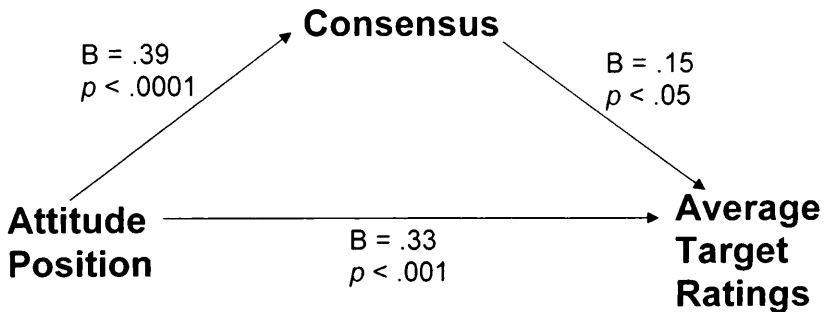Finally, a significant interaction between the target (original target vs.

FIGURE 4. Mediation analysis (Study 4).
Note. Standardized beta weights are reported.

neutral targets) and attitude direction (favorable vs. unfavorable), $F(1, 120) = 9.72$, $p < .003$, showed that the basic correspondence bias effect (i.e., differences in attitudes attributed to favorable and unfavorable targets) was larger when attributions were made to the original than to the neutral targets. As discussed previously, more extreme ratings to the original target can be explained in terms of a heightened tendency to link the target to his or her attitude expressions. Thus, attributions to the original target involve more than attitude transference.

Consensus Estimates and Path Analysis. As in the previous studies, participants exposed to favorable statements estimated that more students favored this position than did participants exposed to unfavorable statements, $F(1, 120) = 21.99$, $p < .0001$. A path analysis was again conducted to assess the direct and indirect influences of attitude position on attributed attitudes. The results shown in Figure 4 demonstrate, as in the previous studies, that attitude position influences attitude attributions both directly and indirectly via its influence on consensus information. These findings again indicate that attitude position independently influences attitude attributions for constrained behavior. Figure 5 repeats this analysis for ratings of the original target. The analysis for the original target shows a direct effect of attitude position but no mediated effect via consensus estimates. The mediated effect for the original target may have been overwhelmed by the extremely strong direct effect of being exposed to favorable or unfavorable attitude statements.

Response-Time Measures. Data from eight participants were discarded because they made a disproportionate number of errors in indicating
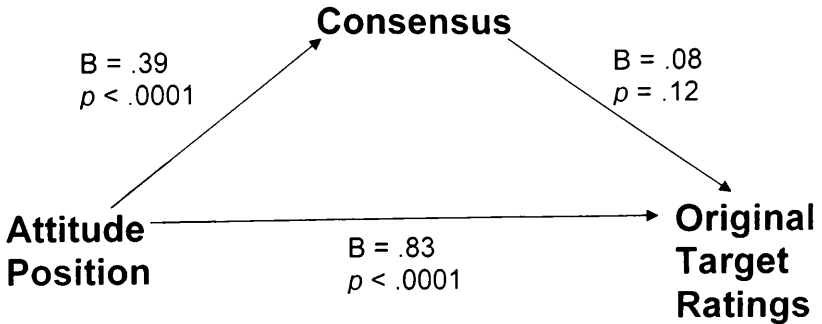
Consensus

B = .39   B = .08
p < .0001   p = .12

**Attitude   Original
Position**   B = .83   **Target
                  p < .0001   Ratings**

FIGURE 5. Mediational analysis for original target (Study 4).
Note. Standardized beta weights are reported.

whether a response was favorable or unfavorable. Trials in which partic-
ipants incorrectly judged a statement were eliminated (4.6% error rate
overall), as were trials in which the RT was higher than three standard
deviations above the mean (errors and outliers were distributed approx-
imately equally across experimental conditions). Finally, the data were
log transformed before the statistical analysis was conducted. The un-
transformed means and standard deviations are displayed in Table 6.

The RT data were submitted to a 2 (attitude direction: favorable or un-
favorable) × 2 (target statement: favorable or unfavorable) mixed
ANOVA with repeated measures on the second factor. As predicted, the
analysis revealed a significant attitude direction × target statement inter-
action, $F(1, 112) = 12.38, p < .001$. The interaction pattern is somewhat ob-
scured, however, by a highly significant main effect showing that
overall, participants responded more quickly to favorable than to unfa-
vorable statements, $F(1, 112) = 19.67, p < .0001$. As Figure 6 shows, when
the general tendency to respond more quickly to favorable statements is
removed (Rosnow & Rosenthal, 1989, 1991), the data clearly indicate
that people exposed to favorable statements responded more quickly to
favorable relative to unfavorable target statements, whereas partici-
pants exposed to unfavorable statements responded more quickly to un-
favorable relative to favorable target statements.

We used our measure of attitude accessibility (response times to favor-
able vs. unfavorable statements) as a mediator of neutral target attribu-
tions. Including this path in the model left the analysis virtually
unchanged. Because the effect of this construct on the dependent vari-

TABLE 6. Mean Response-Times and Standard Deviations in Milliseconds (Study 4)

	Direction of Prime	
	Pro	Anti
Response-Time		
Favorable Target Statements		
M	3662	3822
SD	816	974
n	57	57
Unfavorable Target Statements		
M	4068	3888
SD	877	1084
n	57	57

able was relatively weak ($r = -.15$, $p = .105$), we cannot say definitively whether the failure to find a mediational effect was due to the weakness of the construct or to the actual lack of a mediational role. We address this issue further in the general discussion.

STUDY 5

Study 5 was designed to evaluate another potential contributor to attitude transference. When research participants hear favorable or unfavorable statements expressed on an attitude topic, these statements may bring their own topic-relevant attitudes to the fore. Differences in attitude attributions to neutral targets, therefore, may result from participants using their own attitudes to infer the targets' attitudes. If attitude transference can be completely explained in terms of participants transferring their own attitudes, then attitude transference would be akin to classical projection effects rather than representing a distinct transference phenomenon.

Previous research examining the role of personal attitudes in correspondence bias has produced mixed results. Some studies have reported significant correlations between own and attributed attitudes (Alicke, Zerbst, & LoSchiavo, 1996; Jones & Harris, 1967) whereas others have not (Fleming & Darley, 1989; Jones, Worchel, Goethals, & Grumet, 1971; Miller, 1974; Snyder & Jones, 1974), and one study has obtained a complex interaction with other variables (Jones, Riggs, & Quattrone, 1979). These studies examined the role of participants' preexisting attitudes on their target attributions. However, because research has shown that participants' attitudes sometimes change as a result of hearing or reading the target's views (e.g., Miller, Jones, & Hinkle, 1981; Yandell & Insko, 1977), the argument that participants' attitudes are primed by the
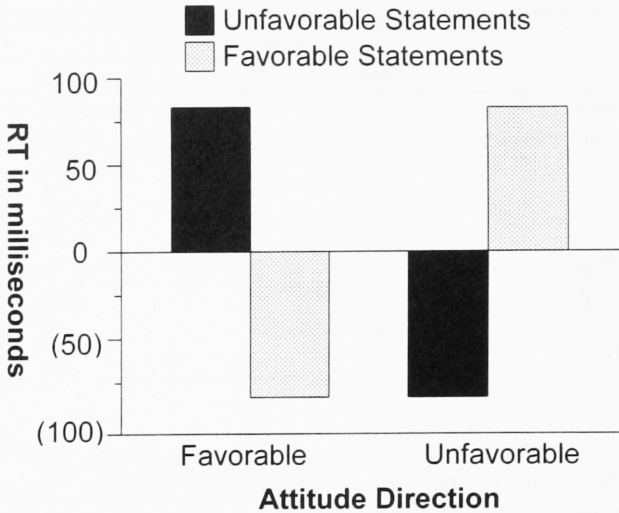
FIGURE 6. Attitude Direction by Target Statement Valence Interaction After Removing the Main Effect of Target Statement Valence (Study 4).

target's statements must account for the possibility that participants hold new attitudes following exposure to the targets' expressions. For this reason, we assessed participants' attitudes after they were exposed to the target's favorable or unfavorable attitude statements. As with consensus estimates, we expected the attitude transference effect to remain significant after accounting for personal attitudes. Showing that the effect of the target's favorable or unfavorable expressions perseveres with personal attitudes and consensus both controlled would provide the strongest demonstration yet of attitude transference as a distinct transference phenomenon.

Another purpose of Study 5 was to assess whether attitude transference effects occur when correspondence bias is attenuated by a constraint. Although manipulation checks in Studies 1 and 2 demonstrated that participants believed the target was more constrained in the constraint than in the free choice conditions, this difference was not reflected in their attitude attributions. Specifically, these studies failed to obtain the usual interaction showing that attitude attributions between participants exposed to favorable versus unfavorable views are less polarized in the constraint conditions than in the free choice conditions. To this point, therefore, we have yet to demonstrate that attitude transference effects occur when attitude attributions to the original target are attenuated by constraints.

The interaction typically obtained in correspondence bias research

could not be assessed in our previous studies because these studies either excluded attributions to the original target (Studies 1–3) or excluded free choice conditions (Studies 3 and 4). Study 5, therefore, included attributions to the original target along with free choice and constraint conditions. Study 5 also included a more powerful constraint than used in the previous studies to enhance the chances of obtaining an interaction between attitude direction (favorable vs. unfavorable) and choice (free choice vs. constraint). Participants in the constraint condition of Study 5 were led to believe that the target had an ulterior motive for expressing favorable or unfavorable views. Research by Fein (Fein, 1996; Fein, Hilton, & Miller, 1990) has shown that correspondence bias is greatly diminished, and often eradicated, when participants are led to believe that targets had ulterior motives for expressing their views. Suspecting an ulterior motive apparently makes people reluctant to accept behavior at face value. By introducing an ulterior motive in the constraint condition, therefore, we expected to obtain the usual interaction between attitude direction and choice on original target attributions. Furthermore, we expected to obtain this same interaction pattern, although somewhat attenuated, for neutral target attributions. This latter finding would show that attitude transference effects occur under both free choice and constraint.

METHOD

Participants. Participants in Study 5 were 186 (74 male and 112 female) undergraduate students whose participation partially fulfilled a course requirement.

Procedure. Study 5 used the same basic methodology as Studies 1 and 2 with two procedural changes. First, participants' attitudes were measured after they were exposed to the target's favorable or unfavorable attitude statements. The second difference was that instead of learning that the target was randomly assigned to express favorable or unfavorable statements, participants in the constraint condition were told the target had an ulterior motive for the statements he wrote.

Upon entering the laboratory, participants were seated at computers and presented with the same instructions as in Study 1. In addition to the information provided in the first study, all participants read the following information:

> The statements you will evaluate today have been written by John, one member of the Scripps team. Before composing the statements, John consulted his advisor to get a few ideas. John always tries to please his advisor

as best he can; he knows that a glowing letter of recommendation will help him land a great job after graduation.

The free choice versus ulterior motive manipulation was introduced at this point. Participants in the free choice condition read the following information:

John's advisor told him that he was free to write any statements that he wanted to write. John, therefore, wrote statements that expressed his own true attitudes.

Participants in the ulterior motive condition read the following information:

John's advisor told him that he was free to write any statements that he wanted to write. His advisor made it perfectly clear, however, that he fully supports (opposes) course standardization. After giving the situation some thought, John decided that it would be best if he wrote statements that support (oppose) course standardization.

After reading the 15 positive or negative statements, filling out the manipulation check items, and rating the original and neutral targets, participants were asked to rate their personal attitudes. Participants indicated their attitudes toward course standardization by placing an "X" on a 147mm. line which was anchored by "strongly oppose" and "strongly favor."

RESULTS AND DISCUSSION

Manipulation Checks. The first manipulation check assessed whether participants were aware of the favorable versus unfavorable nature of the propositions. This manipulation check was successful, $F(1, 182) = 496.06, p < .0001$, showing that participants exposed to favorable propositions considered those propositions to be more favorable ($M = 8.09, SD = 1.32$) than did participants exposed to unfavorable propositions ($M = 2.61, SD = 1.97$).

The second manipulation check assessed whether participants recognized that the attitude propositions they read were freely expressed or constrained by ulterior motives. This difference was significant, $F(1, 182) = 118.76, p < .0001$, showing that participants in the free choice condition perceived the original target to be freer to express his true views ($M = 8.15, SD = 1.25$) than did participants in the ulterior motive condition ($M = 5.22, SD = 2.25$).

TABLE 7. Mean Attributions and Standard Deviations for Original and Neutral Targets' Beliefs About College Course Standardization (Study 5)

| | Attitude Expressed | | | |
| | Free Choice | | Ulterior Motive | |
	Pro	Anti	Pro	Anti
Attitude Attribution				
Original Target				
M	8.20	2.04	6.59	4.19
SD	1.13	1.41	1.44	1.59
n	44	45	49	48
Neutral Targets				
M	6.19	5.14	5.64	5.17
SD	0.92	0.68	0.90	0.73
n	44	45	49	48

Attitude Attributions. Table 7 displays means and standard deviations for attitude attributions made to both the original and neutral targets. Original and neutral targets were included as a within-subjects factor in an analysis of variance along with the between-subjects factors of attitude direction (favorable or unfavorable) and level of choice (free choice or ulterior motive). Regarding the original target, results showed that participants exposed to favorable statements ascribed more favorable attitudes to the writer ($M = 7.35$, $SD = 1.53$) than did those exposed to unfavorable statements ($M = 3.15$, $SD = 1.85$), $F(1, 182) = 427.60$, $p < .0001$. This effect was qualified, however, by a significant attitude direction by level of choice interaction, $F(1, 182) = 82.23$, $p < .0001$. Furthermore, although the differences were attenuated in the ulterior motive condition, simple effects tests revealed that the effect of attitude direction (favorable or unfavorable) was significant in both the free choice condition, $F(1, 182) = 424.18$, $p < .0001$, and in the ulterior motive condition, $F(1, 182) = 70.43$, $p < .0001$. This interaction pattern replicates the usual interaction between choice and attitude direction obtained in most correspondence bias studies and also shows the basic correspondence bias effect, namely, a difference between attributed attitudes to favorable versus unfavorable targets in the constraint condition.

The same analyses were significant for neutral targets. Participants exposed to favorable statements ascribed more favorable attitudes to the neutral targets ($M = 5.90$, $SD = 0.95$) than did those exposed to unfavorable statements ($M = 5.15$, $SD = 0.70$), $F(1, 182) = 40.49$, $p < .0001$. Consistent with analyses involving the original target, this effect was qualified by a significant attitude direction by level of choice interaction, $F(1, 182) = 5.94$, $p < .02$. This interaction followed the same pattern as for the origi-

nal target, that is, the effects of attitude direction were smaller in the ulterior motive condition than in the free choice condition. As with the original target analysis, simple effects tests revealed that the effect of attitude direction (favorable or unfavorable) was significant in both the free choice condition, $F(1, 182) = 37.13, p < .0001$, and in the ulterior motive condition, $F(1, 182) = 8.05, p < .01$. These results show that participants' correction for the constraint diminishes the effects of attitude direction for ratings of the neutral targets as well as for the original target. This finding suggests that attitude transference is a residual effect that remains even when corrections are made for the original communication's validity. In other words, the knowledge that the original target's views were randomly assigned leads people largely to discount that view. The original communication's validity is a factor in its transferability as evidenced by the diminished tendency to ascribe that view to neutral targets. Perceived invalidity of the original communication is insufficient, however, to completely negate attitude transference as shown by the residual effect on neutral target attributions. The fact that attitude transference effects remain even when attributions are tempered by the constraint demonstrates the robustness of the effect.

Finally, a significant 3-way interaction among the target (original target or neutral targets), attitude direction (favorable or unfavorable), and level of choice (free choice or ulterior motive), $F(1, 182) = 53.22, p < .0001$, demonstrated that the basic correspondence bias effect was larger when attributions were made to the original than to the neutral targets. This finding replicates the results of Study 4 and indicates that attributions to the original target are influenced by the fact that the original target, in contrast to the neutral targets, actually expressed an attitude.

Personal Attitudes, Consensus Estimates, and Path Analysis. Each of our previous studies showed that exposing participants to one side of an attitude issue influences their consensus estimates. Furthermore, research on the false consensus effect (Ross, Greene, & House, 1977) shows that personal attitudes predict consensus estimates (Alicke & Largo, 1995). Based on these findings, we assessed the possibility that personal attitudes mediate the effect of attitude direction on consensus estimates.

We assumed that the direct effect of attitude direction on neutral target attributions would remain after accounting for both personal attitudes and consensus estimates. A path analysis was conducted to determine the tenability of this assumption. The results of the path analysis are shown in Figure 7. In addition, Table 8 displays means and standard deviations for participants' personal attitudes and consensus estimates. Consistent with the foregoing reasoning, the path analysis first shows that the effect of attitude direction on consensus estimates is mediated by participants' personal attitudes. The path analysis further
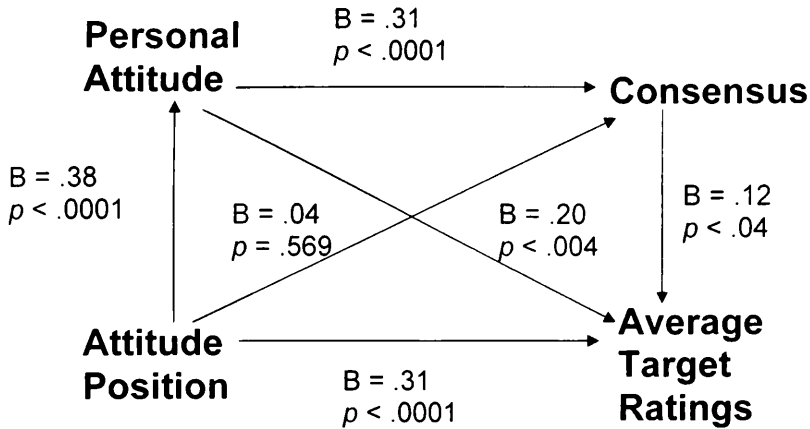
Personal Attitude   B = .31
                         $p < .0001$ → **Consensus**

B = .38
$p < .0001$ B = .04   B = .20   B = .12
                  $p = .569$ $p < .004$ $p < .04$

Attitude Position   →**Average Target Ratings**
                         B = .31
                         $p < .0001$

FIGURE 7. Mediation analysis (Study 5).

Note. Standardized beta weights are reported.

shows that personal attitudes have both direct and indirect (via consensus) effects on neutral target ratings. Most importantly for present purposes, however, was the finding that exposing participants to favorable or unfavorable attitudes directly influenced attitude attributions to neutral targets when both personal attitudes and consensus estimations were taken into account. Thus, the attitude transference phenomenon cannot be fully explained by heightened consensus perceptions or transference of personal attitudes.

GENERAL DISCUSSION

Cognitive transference is a general term describing the tendency for people to transfer thoughts, motives, beliefs and attitudes from one source to others. Classical and social projection, which dominate the empirical literature on this topic, assume that self-involvement underlies cognitive transference effects. The attitude transference phenomenon demonstrated in the present five studies, however, show that both self-related and non-self-related mental contents are transferable. In contrast to priming, which can also be construed as a type of non-self-related transference, attitude transference involves taking attitude statements generated by one source and attributing them to another. In most priming studies, an externally primed construct is used as a guide for interpreting ambiguous behavior. In this regard, attitude transference is formally closer to social projection than to priming although it differs from social projection in two important ways: first, in that the source of the gener-

TABLE 8. Mean Personal Attitude Ratings, Consensus Estimates, and Standard
Deviations regarding College Course Standardization (Study 5)

	Attitude Expressed	
	Pro	Anti
Personal Attitude		
M	94.19	65.44
SD	35.79	35.04
n	93	93
Consensus Estimates		
M	52.42	49.37
SD	9.28	9.44
n	93	93

Note. For each variable, means differ significantly at $p < .05$.

ated attitude is someone other than the self, and second, in that the re-
sponse measure involves attributions to specific targets rather than
consensus estimates.

A secondary purpose of this research was to suggest how cognitive
transference might be involved in correspondence bias. During the first
phase of most attitude attribution experiments, participants are exposed
to targets who express favorable or unfavorable views on controversial
issues. According to the cognitive transference scheme presented in Ta-
ble 1, these statements generate positive or negative attitude positions
which are then transferred back to the person who generated them, de-
spite the fact that the positions were randomly assigned. In correspon-
dence bias, as in the trait transference phenomenon studied by
Skowronski et al. (1998), transference is reflexive in that the contents are
transferred back to the target who generated them. We reasoned that if
attitudes generated by the original target are transferred to neutral tar-
gets, it seems reasonable to assume that some portion of the correspon-
dence bias effect on the original target is also due to attitude
transference. In essence, we used attitude transference to neutral targets
as a basis for arguing that a similar process characterizes reflexive trans-
ference to the original target.

Attitude transference to neutral targets was demonstrated in five sep-
arate studies: In the first study, participants read 15 target-generated
statements that were either favorable or unfavorable toward course
standardization. Replicating the usual correspondence bias effect, par-
ticipants who read favorable statements ascribed more favorable atti-
tudes than those who read unfavorable statements, although they knew
the statements were randomly assigned. The difference between this

and previous attitude attribution studies, however, was that attitudes were transferred to neutral targets whose images were conveyed on a computer screen rather than to the target who actually expressed the attitudes.

The second and third studies were conceptually similar to the first, except that the attitude dimension involved homosexuals participating in the military and attitudes were expressed on videotape rather than on computer. Again, attitude transference was obtained for attributions of neutral targets who appeared via still videotaped images. Study 3 showed that these attitude transference effects persevered for a week after the original manipulation, thereby demonstrating the robustness of the effect.

The fourth study measured response times to assess whether favorable and unfavorable attitude positions were made differentially accessible as a result of reading the targets' statements. In addition to replicating the attitude transference effect obtained in the previous studies, the results of Study 4 showed that participants exposed to favorable statements responded more quickly to favorable than to unfavorable statements relative to those exposed to unfavorable statements. Study 4, therefore, provided experimental evidence that favorable and unfavorable attitudes were made differentially accessible by the original target's attitude expression. However, the absence of a mediating effect of attitude accessibility on attitude attributions raises questions as to whether heightened accessibility is involved in attitude transference.

Each of the first four studies also assessed the possibility that hearing favorable or unfavorable attitude statements influences participants' estimates of the percentage of their peers who hold favorable or unfavorable attitudes. If so, heightened consensus might provide a sufficient explanation of attitude transference effects. Our data revealed that consensus estimates did consistently influence attitude attributions. However, each study also showed that attitude transference effects persevered while accounting for the influence of perceived consensus. Thus, while consensus estimates play a role in attitude attributions to neutral targets, they cannot completely explain attitude transference.

Study 5 was designed to assess the possibility that participants' own views are made more accessible as a result of hearing the target's attitude statements, and that participants transfer their own views—rather than the views expressed by the original target—onto the neutral targets. To assess this possibility, we measured both personal attitudes and consensus estimates after participants were exposed to the target's statements. As with consensus estimates, these analyses showed that personal attitudes were significantly related to neutral target attributions.

However, the fact that the effects of being exposed to favorable or unfavorable attitudes persevered after covarying the effects of both personal attitudes and consensus estimates suggests that other factors contribute to attitude transference.

In contrast to the first four studies, Study 5 incorporated all components of the typical correspondence bias paradigm, namely, free choice and constraint conditions as well as attributions to the original target. Because the previous studies either excluded original target attributions or did not include free choice conditions, it was impossible to assess the usual interaction obtained in correspondence bias studies between attitude direction (favorable vs. unfavorable) and choice (free choice vs. constraint). The failure in these studies to obtain a difference between the free choice and constraint conditions on neutral target ratings raised questions about the present findings' applicability to correspondence bias. Specifically, in the absence of this interaction, we could not assess whether attitude transference onto neutral targets occurs when attributions to the original target are attenuated by the constraint. Study 5 resolved this ambiguity, however, by demonstrating the traditional finding that attitude attributions to the original target were more polarized under free choice than under constraint.

The fact that the same interaction between attitude direction and choice occurred on neutral target ratings as on original target ratings suggests that neutral target attributions are influenced by the perceived validity of the communication. In other words, this interaction shows that participants were sensitive to the fact that constrained messages are less valid indicators of the original target's true position and showed a diminished tendency to transfer such messages. Nevertheless, the residual transference effect was still significant in the constrained condition. These findings raise interesting questions about the relative roles of perceived validity and accessibility in attitude attributions. The constraint condition in correspondence bias studies is similar to hearing a communication from a low credible source. In both instances, a discounting cue (experimental assignment or the nature of the source) leads observers to downplay the message's significance. As research on the sleeper effect (e.g., Cook & Flay, 1978; Cook et al., 1979; Gillig & Greenwald, 1974; Pratkanis, et al., 1988) shows, discounted messages can alter attitudes, although a delay is typically required for the message and the message source to be dissociated. Our studies differed from those in sleeper effect research in that attitude attributions were made for neutral targets rather than for oneself, and no delay was required to obtain attitude transference effects. The parallel between these two lines of research lies in the fact that being exposed to a message influences attitude attribu-

tions although the validity of the message source is discounted. As the results of Study 5 suggest, discounting the source of the original message attenuates the extent to which the message is attributed to neutral targets although the message's accessibility results in a residual transference effect. In other words, the source's perceived invalidity only partly overrides attitude transference. This raises the question for future research of what types and degrees of perceived invalidity are sufficient to counteract attitude transference.

Further research is needed to clarify the mechanisms underlying the attitude transference effect. These present studies clearly show that being exposed to another person's attitude heightens perceived consensus for that position and also alters some participants' personal attitudes. Both consensus estimates and personal attitudes mediated neutral target attributions. In addition, exposure to an attitude position heightened the cognitive availability of that position. Cognitive availability is another factor that may account for attitude transference effects. The availability or accessibility explanation assumes that when people seek evidence upon which to base their target judgments they tend to access the most available relevant information. In trying to estimate a person's attitude, therefore, people call to mind the first attitude positions they can think of. This interpretation is closest to cognitive availability interpretations of social projection (e.g., Marks & Miller, 1987; Sherman, et. al., 1983). However, whereas the cognitive availability view of social projection assumes that people project their preferences because they selectively recall instances in which others have agreed with them, the present view simply assumes that people transfer the most available preferences to others. Although we did not obtain a mediating effect of accessibility on neutral target attributions, this nonfinding could possibly have been due to an inadequate accessibility measure. Further research is needed, therefore, to assess the accessibility hypothesis with other accessibility measures.

In future research, a more convincing link between accessibility and correspondence bias might be established by varying the accessibility of the original target's attitude expressions. To date, accessibility manipulations in correspondence bias have primarily entailed increasing the constraint's salience (e.g., Miller, Jones, & Hinkle, 1981). As Gilbert and Jones (1986) have shown, however, even people who are responsible for constraining the target's behavior attribute correspondent attitudes; thus, correspondence bias occurs even when the constraint is extremely obvious. Accessibility in relation to attitude transference refers to the availability of the attitude position expressed rather than to the constraint. Accessibility in this context could be manipulated by creating a

delay between the target's presentation and the attribution of attitudes to the target, or by varying the number of arguments the target presents. If reducing cognitive availability reduces correspondence bias to both the original and to neutral targets, this would more firmly establish cognitive availability as an explanation for correspondence bias.

An unresolved issue in these studies involves the specific contents transferred. We presented participants in each study with a series of attitude propositions designed to represent the favorable or unfavorable end of the attitude continuum and then measured their general attitudes toward the neutral targets. We cannot be sure, therefore, whether participants transferred the specific propositions participants heard or the general attitude underlying these propositions. This raises a more general question regarding the types of transferrable mental contents. Psychodynamically oriented research studies the projectibility of specific thoughts, motives, and traits, whereas social projection research, which occupies the largest tract of the projection landscape, studies attitude and behavior projection. Some of the material projected in previous research has been strongly valenced, whereas other projections entail mundane preferences such as those for lima beans. In terms of projecting self-related contents, even mundane preferences may be relatively available. However, projection of non-self-related contents may require more strongly valenced material. We are uncertain, for example, whether hearing another person express a preference for lima beans would lead to lima bean projections onto neutral targets, but the answer to this question awaits future research on attitude transference

REFERENCES

Alicke, M. D., & Largo, E. (1995). The role of the self in the false consensus effect. *Journal of Experimental Social Psychology, 31*, 28–47.

Alicke, M. D., Zerbst, J. I., & LoSchiavo, F. M. (1996). Personal attitudes, constraint magnitude, and correspondence bias. *Basic and Applied Social Psychology, 18*, 211–228.

Cook, T. D., & Flay, B. R. (1978). The persistence of experimentally induced attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 1–57).

Cook, T. D., Gruder, C. L., Hennigan, K. M., & Flay, B. R. (1979). History of the sleeper effect: Some logical pitfalls in accepting the null hypothesis. *Psychological Bulletin, 86*, 662–679.

D'Agostino, P. R., & Hawk, M. (1998). The transfer of actor–trait associations inferred from behavior. *Social Cognition 16*, 391–399.

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology, 70*, 1164–1184.

Fein, S., Hilton, J. L., & Miller, D. T. (1990). Suspicion of ulterior motives and the correspondence bias. *Journal of Personality and Social Psychology, 58*, 753–764.

Freud, S. (1956). Further remarks on the defense neuro-psychoses. In *Collected papers of Sigmund Freud* (Vol. 1). London: Hogarth Press.

Gilbert, D. T., & Jones, E. E. (1986). Perceiver-induced constraint: Interpretations of self-generated reality. *Journal of Personality and Social Psychology, 50*, 269–280.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21–38.

Gillig, P. M., & Greenwald, A. G. (1974). Is it time to lay the sleeper effect to rest? *Journal of Personality and Social Psychology, 29*, 132–139.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*, 141–154.

Holmes, D. S. (1968). Dimensions of projection. *Psychological Bulletin, 69*, 248–268.

Horney, K. (1939). *New ways in psychoanalysis.* New York: Norton.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1–24.

Jones, E. E., Riggs, J. M., & Quattrone, G. (1979). Observer bias in the attitude attribution paradigm: Effect of time and information order. *Journal of Personality and Social Psychology, 37*, 1230–1238.

Jones, E. E., Worchel, S., Goethals, G. R., & Grumet, J. (1971). Prior expectancy and behavioral extremity as determinants of attitude attribution. *Journal of Experimental Social Psychology, 7*, 59–80.

Katz, D., & Allport, F. (1931). *Students attitudes.* Syracuse: Craftsfman Press.

Krueger, J. (1998). On the perception of social consensus. In M. Zanna (Ed.). *Advances in experimental social psychology* (Vol. 30, pp. 163–240.). Hillsdale, NJ: Erlbaum.

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin, 102*, 72–90.

Miller, A. G. (1974). Perceived freedom and the attribution of attitudes. *Representative Research in Social Psychology, 5*, 61–80.

Miller, A. G., Ashton, B., & Mishal, M. (1990). Beliefs concerning the features of constrained behavior: A basis for the fundamental attribution error. *Journal of Personality and Social Psychology, 59*, 635–650.

Miller, A. G., Jones, E. E., & Hinkle, S. (1981). A robust attribution error in the personality domain. *Journal of Experimental Social Psychology, 17*, 587–600.

Murray, H. A. (1933). The effect of fear upon estimates of the maliciousness of other personalities. *Journal of Abnormal and Social Psychology, 4*, 310–339.

Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead. Long live the sleeper effect. *Journal of Personality and Social Psychology, 54*, 203–218.

Rosnow, R. L., & Rosenthal, R. (1989). Definition and interpretation of interaction effects. *Psychological Bulletin, 105*, 143–146.

Rosnow, R. L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at only the interaction (unless all main effects are zero). *Psychological Bulletin, 110*, 574–576.

Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301.

Sherman, S. J., Presson, C. C., Chassin, L., Corty, E., & Olshavsky, R. (1983). The false consensus effect in estimates of smoking prevalence. *Personality and Social Psychology Bulletin, 9*, 197–207.

Sherman, S. J., Presson, C. C., & Chassin, L. (1984). Mechanisms underlying the false consensus effect: The special role of threats to self. *Personality and Social Psychology Bulletin, 10*, 127–138.

Skowronski, J. J., Carlson, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait trans-

ference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology, 74,* 837–848.

Snyder, M. L., & Jones, E. E. (1974). Attitude attribution when behavior is constrained. *Journal of Experimental Social Psychology, 10,* 857–864.

Yandell, B., & Insko, C. A. (1977). Attribution of attitudes to speakers and listeners under assigned-behavior conditions: Does behavior engulf the field? *Journal of Experimental Social Psychology, 3,* 269–278.

Positive Expectations Encourage Generalization From a Positive Intergroup Interaction to Outgroup Attitudes

Matthew P. Deegan[1], Eric Hehman[2], Samuel L. Gaertner[1], and John F. Dovidio[3]

Abstract

The current research reveals that while positive expectations about an anticipated intergroup interaction encourage generalization of positive contact to outgroup attitudes, negative expectations restrict the effects of contact on outgroup attitudes. In Study 1, when Blacks and Whites interacted with positive expectations, interaction quality predicted outgroup attitudes to a greater degree than when groups interacted with negative expectations. When expectations (Studies 2 and 3) and the actual interaction quality (Study 4) were manipulated orthogonally, negative expectations about the interaction predicted negative outgroup attitudes, regardless of actual interaction quality. By contrast, participants holding positive expectations who experienced a positive interaction expressed positive outgroup attitudes, whereas when they experienced a negative interaction, they expressed outgroup attitudes as negative as those with negative expectations. Across all four studies, positive expectations encouraged developing outgroup attitudes consistent with interaction quality.

Positive intergroup contact generally produces improved intergroup attitudes, even when the conditions for contact specified by contact theory (e.g., equal status, common fate; Allport, 1954) are not fully satisfied (Pettigrew & Tropp, 2006; Pettigrew, Tropp, Wagner, & Christ, 2011). This effect occurs not only for attitudes toward outgroup members directly involved in the contact situation but also generalizes to the outgroup category. However, intergroup contact is not uniformly effective for improving intergroup relations (Pettigrew & Tropp, 2006). The present research investigated one factor, expectations about the intergroup interaction, that can significantly affect the nature of intergroup contact and the impact of that contact on outgroup attitudes.

Expectations profoundly shape intergroup interactions, and thus facilitating more positive expectations can improve interactions (Mallett & Wilson, 2010; Plant, Butz, & Tartakovsky, 2008). Positive expectations reduce intergroup anxiety (Riek, Mania, & Gaertner, 2006). In turn, reduced anxiety can decrease outgroup avoidance, increase desire to interact with outgroup members, and decrease feelings of outgroup hostility (Plant & Devine, 2003). Indeed, improving expectations of interracial interaction can increase the likelihood of developing new cross-race friendships (Mallett

& Wilson, 2010), which, in turn, improves intergroup attitudes (Pettigrew, 1998).

However, negative expectations may play an even more prominent role, undermining the effectiveness of intergroup contact for improving intergroup attitudes. Understanding the influence of expectations on intergroup interactions and outcomes is important because people typically approach intergroup interactions with trepidation and anxiety (Dovidio, Kawakami, & Gaertner, 2002; Mallett, Wilson, & Gilbert, 2008; Shelton & Richeson, 2005). These negative expectations can arouse negative emotions, which shape intergroup attributions and behavior (Plant et al., 2008). Negative expectations produce biases in how people encode outgroup members' behaviors (Maass, Milesi, Zabbini, & Stahlberg,

[1]University of Delaware, Newark, USA
[2]New York University, New York City, USA
[3]Yale University, New Haven, CT, USA

Corresponding Author:
Matthew P. Deegan, Department of Psychological and Brain Sciences, University of Delaware, 105 The Green, Room 108, Newark, DE 19716, USA.
Email: mdeegan@psych.udel.edu

1995; Rothbart & John, 1985). For example, negative outgroup behaviors, which are generally consistent with stereotype-based expectations, are encoded more abstractly and with an internal locus of causality (e.g., Jim is stupid). Comparatively, positive outgroup behaviors (e.g., Jim did well on his test), which are usually inconsistent with stereotype-based expectations, are encoded more concretely and with an external locus of causality. Because people seek to confirm rather than disconfirm their expectations, disconfirmation of negative expectations is less likely (Snyder, Tanke, & Berscheid, 1977).

How people process information about their intergroup encounters is particularly important because experiences in intergroup interactions often differ from expectations. For instance, Shelton and Richeson (2005) found that both Whites and Blacks avoid interracial contact because they anticipated rejection of their positive overtures by members of the other group. In reality, both groups welcomed such overtures. While prior research has established the direct impact of expectations on intergroup attitudes and behaviors, the present work investigated the way expectations for intergroup contact moderate the effect of a positive or negaitve intergroup interaction on general outgroup attitudes.

In general, people regard negative information as more diagnostic than positive information about others' character because negative behaviors are typically counter-normative and thus elicit more confident dispositional attributions than positive behaviors (Jones & Davis, 1965; Ybarra, Schaberg, & Keiper, 1999). Therefore, people are less motivated to examine the veracity of negative information and more motivated to test positive expectations in subsequent interactions (Ybarra et al., 1999). For example, after forming positive or negative expectations of an individual, people recalled expectancy-inconsistent behaviors better when they anticipated the interaction to be positive (Ybarra et al., 1999). Specifically, those with a positive expectancy remembered more negative information compared with the amount of positive information remembered by those with negative expectations. Also, when expectations were positive, memory for all traits or behaviors, whether positive, negative, or neutral, was better than when expectations were negative (Ybarra et al., 1999). Thus, positive expectations enhance the use of online, objective information verification when forming an impression of a person, such that positive expectations lead to more accurate impressions (Ybarra et al., 1999).

Similarly, it takes longer to make dispositional judgments about favorable compared with unfavorable targets (Lingle & Ostrom, 1979), and people request more information before making decisions about favorable compared with unfavorable targets (Yzerbyt & Leyens, 1991). In addition, positive expectations lead to more engagement in the task of evaluating others (Ybarra et al., 1999). Indeed, it takes more frequent observations of positive behavior to disconfirm negative beliefs than for negative behavior to disconfirm positive beliefs (Rothbart & John, 1985). Thus, with positive expectations, individuals tend to be more engaged, pay more attention, and test their positive expectations more robustly, compared with negative expectations (Ybarra et al., 1999). By contrast, negative expectations prevent people from forming accurate impressions of others, despite the availability of positive information, because people tend to form negative evaluations more quickly, and they are less motivated to test these evaluations, leading to shallower, less careful information processing about individuals (Ybarra, 2002).

The present research extends previous work on interpersonal expectations to the intergroup context, examining the unique role expectations play in influencing intergroup attitudes. Whereas previous research shows that expectations influence immediate interactions in expectation-confirming ways (Snyder et al., 1977), our perspective suggests that the valence of expectations has a more nuanced effect on the generalizability of intergroup contact. Given that positive social expectations generally motivate a careful evaluation of one's beliefs about an individual (Ybarra et al., 1999) and negative expectations produce a narrow focus on negative information (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Gotlib et al., 2004), we proposed that an analogous process may occur at the intergroup level. Specifically, we hypothesized that positive expectations increase sensitivity to both positive *and* negative aspects of an intergroup interaction, leaving participants open to developing either more favorable or more pejorative outgroup attitudes, whereas negative expectations render outgroup attitudes relatively resistant to change.

Study 1 utilized a correlational approach that arranged for cross-race interactions to assess the effects of expectations and evaluations of the interaction on outgroup category attitudes. Study 2 extended this work by experimentally manipulating expectations of an intergroup interaction involving Republicans and Democrats to be positive or negative. Study 3 replicated the previous studies using minimal groups while employing a yoked experimental design intended to permit those with positive and negative expectations to observe the same actual interaction. Finally, Study 4 involved minimal groups and independently manipulated both the expectations and the quality of the intergroup interaction using trained confederates.

Study 1

Study 1, using a correlational approach, examined how expectations and perceived interaction quality jointly predict outgroup attitudes. Same-sex groups of White and Black participants met together to complete a common task. We assessed expectations for the interaction, perceived interaction quality, and outgroup attitudes. Based on previous work on expectations for intergroup interactions (see West & Dovidio, 2012, for a review), we predicted that expectations would have a direct effect on perceived interaction quality because people tend to approach interactions

in expectancy-confirming ways. In addition, our novel primary hypothesis was that expectation valence would moderate the effect of the experienced interaction quality on outgroup attitudes. In particular, because negative expectations encourage a rigid focus on negative information and positive expectations facilitate processing a wider range of information (e.g., Baumeister et al., 2001), we predicted that perceived interaction quality would more strongly predict outgroup attitudes for those with positive expectations.

Method

Participants and design. Sixty-four Black and White undergraduates (40 women) participated for partial course credit met in same-sex sessions (*n* = 16) involving four participants (i.e., two 2-person groups).

Procedure. Following a procedure adapted from Gaertner, Mann, Murrell, and Dovidio (1989), individual participants first met and interacted in small racially homogeneous two-person groups to increase the salience of group membership prior to the four-person intergroup interaction. Thus, Black dyads and White dyads initially reported to different locations and were greeted by different experimenters. Then, both dyads entered a larger room where experimenters "completed paperwork" as participants entered. This initial meeting was intended to alert participants to the race of the members of the other dyad. After approximately 10 s, experimenters led each dyad into smaller rooms for their separate intraracial interaction to further increase the salience of their racial group memberships.

During this initial interaction, dyads worked on a Winter Survival (Johnson & Johnson, 1975) task for 6 min during which participants are asked to rank order a number of items salvaged from a plane crash in terms of their importance to survival. Then, participants were told they would interact with the other-race dyad they met earlier and asked to report the race of the other group and their expectations for the upcoming interaction on a 1 to 7 scale from *not at all* to *very much* (nine items, five reverse-scored: I expect the interaction in the next room to be Cooperative, Tense [R], Friendly, Competitive [R], Quarrelsome [R], Unpleasant [R], Agreeable, Successful, Annoying [R]; α = .82).

Both groups then entered a larger room and worked together to reach a consensus solution to the Winter Survival task. As motivation, participants were promised that the four-person group with the best solution over the semester would receive US$100.00. Groups were required to reach a consensus within 10 min, and experimenters updated the groups of the time remaining throughout the session. Afterward, participants completed a questionnaire asking about perceived interaction quality (nine items, five reverse-scored: Please describe your impression of the interaction [1 = *not at all* to 7 = *very much*]: Cooperative, Tense [R], Friendly, Competitive [R], Quarrelsome [R],

Unpleasant [R], Agreeable, Successful, Annoying [R]; α = .85). Participants then made group-level evaluations of the outgroup (four items, two reverse-scored: When thinking of [White or Black] people *in general*, how much do you see them as [1 = *not at all* to 7 = *very much*]: Cooperative, Aggressive [R], Friendly, Suspicious [R]; α = .72).

Results

Nine participants who incorrectly identified the race of the group they expected to interact with were removed from analysis (remaining *N* = 55). These data were analyzed using hierarchical linear modeling (HLM) using HLM 6.06 (Raudenbush, Bryk, & Congdon, 2008). Data were collected from participants in four-person sessions and are therefore interdependent. HLM accounts for this shared variance and generates parameter estimates less biased than traditional measures. HLM coefficients are interpreted similar to unstandardized beta weights. In our analyses, we allowed only Level 2 intercepts to vary. For all models, participant race (Black or White) was entered at Level 1 and gender was entered at Level 2. In all models for Study 1, neither participant race nor participant gender produced significant main or interaction effects and were therefore removed from each model, unless explicitly mentioned. The inclusion of these predictors did not alter results.

Expectations and quality. To examine whether perceived interaction quality was influenced by expectations, perceived interaction quality was regressed on the model described above, with the addition of grand-mean centered expected interaction quality as a Level 1 predictor. This model explained 14% of the Level 1 variance in perceived interaction quality.[1] As predicted, expected interaction quality predicted perceived interaction quality (γ = 0.335, *SE* = 0.123, *p* = .010, 95% confidence interval [CI] = [0.094, 0.576]), suggesting that participants' expectations had a direct effect on their perceptions of how well the interaction proceeded. Participants with more favorable expectations of the interaction reported more positive interaction quality. Overall, participants experienced a positive interaction (*M* = 6.42, *SD* = 0.623) on a 1 to 7 scale.

Outgroup attitudes. We predicted that for individuals expecting a more positive interaction, evaluations of the outgroup category would be more sensitive to their perceptions of interaction quality. Thus, perceived interaction quality should predict outgroup attitudes to a greater degree for participants holding positive than negative expectations. We regressed outgroup category attitudes on the model described above, with the addition of grand-mean centered expected interaction quality, perceived interaction quality, and their multiplicative interaction term as Level 1 predictors. This model explained 17% of the Level 1 variance and 95% of the Level 2 variance in outgroup attitudes. Results supported our hypothesis.
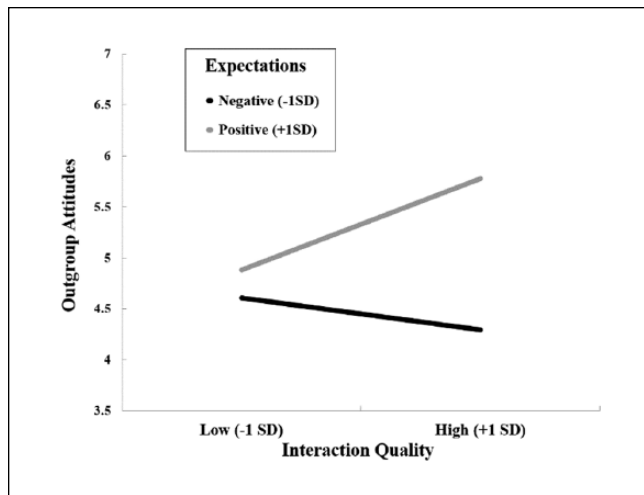
Figure 1. Expected quality and perceived quality interact to predict outgroup attitudes for Black and White participants.

We found a significant main effect of expected interaction quality ($\gamma = 0.623$, $SE = 0.189$, $p = .002$, 95% CI = [0.253, 0.993]), suggesting that more positive expectations elicited more favorable outgroup attitudes. Although we found no main effect of perceived interaction quality on outgroup attitudes ($\gamma = 0.235$, $SE = 0.176$, $p = .190$, 95% CI = [−0.110, 0.580]), we did unexpectedly find an interaction with participant gender ($\gamma = 0.692$, $SE = 0.262$, $p = .013$, 95% CI = [0.178, 1.21]). Simple slopes revealed that for males, perceived interaction quality predicted outgroup attitudes ($\gamma = 0.928$, $SE = 0.194$, $t = 4.78$, $p < .001$); no such effect was present for females ($\gamma = 0.236$, $SE = 0.176$, $t = 1.34$, $p = .190$). As predicted, we found a reliable interaction between expected interaction quality and perceived interaction quality ($\gamma = 0.685$, $SE = 0.169$, $p < .001$, 95% CI = [0.354, 1.02]). This statistically significant interaction is of the most theoretical significance because it represents a test of our central hypothesis that, for cross-race interactions, positive expectations will lead to more positive outgroup evaluations than negative expectations (see Figure 1).

As predicted, decomposing the interaction revealed that for participants with *positive* expectations (+1 *SD*), perceived interaction quality predicted outgroup attitudes ($\gamma = 0.723$, $SE = 0.272$, $t = 2.66$, $p = .012$). However, participants with relatively negative expectations for the interaction (−1 *SD*) showed a marginally significant negative relationship between perceived interaction quality and outgroup attitudes ($\gamma = -0.252$, $SE = 0.131$, $t = -1.92$, $p = .065$), suggesting that negative expectations may have inhibited any positive effect of outgroup contact (see Figure 1).

We originally predicted that participants with positive expectations for intergroup interactions would be sensitive to both positive and negative aspects of the intergroup interaction. Our results, however, suggest that for participants who

perceived their interactions to be relatively negative, positive expectations did not lead to worse evaluations of the outgroup compared with negative expectations. For participants who perceived the interaction to proceed negatively (−1 *SD*), there was no difference between participants with positive (+1 *SD*) or negative expectations (−1 *SD*; $\gamma = 0.196$, $SE = 0.182$, $t = 1.08$, $p = .289$). Therefore, positive expectations led the participants to generalize their positive, but not negative, contact experiences, whereas negative expectancies limited the generalization of positive experiences.

Discussion

The overall pattern of results is consistent with our predictions. First, as expected, for both Blacks and Whites, participants with more positive expectations experienced their intergroup encounter more positively. Second, the impact of experienced interaction quality on outgroup attitudes was moderated by whether the initial expectations of the interracial interaction were positive or negative. Specifically, interaction quality predicted outgroup attitudes for those with positive expectations but not for those with negative expectations. This result supports our hypothesis that negative expectations would limit generalization from positive contact to outgroup attitudes. Notably, participants generally had a positive contact experience ($M = 6.42$, $SD = 0.623$). Despite the generally positive contact participants experienced, those approaching the interaction with relatively negative expectations did not generalize this contact from perceived interaction quality to outgroup attitudes.

Study 1 thus demonstrated support for the two hypothesized ways that expectations can affect intergroup relations—both initially in predicting interactions consistent with the expectations and also in moderating the generalizability of the effects of contact, particularly the benefits of positive contact, on outgroup attitudes. However, we acknowledge that even though expectations, perceived quality of contact, and outgroup attitudes were assessed sequentially across time, causal inferences are limited by the correlational design. More definitive evidence for the hypothesized dynamics requires a direct manipulation of intergroup expectations. Study 2 pursues this approach and considers a different domain of intergroup relations involving political affiliation rather than race.

Study 2

In Study 2, a two-person group of Democrats and a two-person group of Republicans participated in an intergroup interaction in a two-level (expectations: negative, positive) design. In contrast to the correlational approach of Study 1, we manipulated expectations about an intergroup interaction by providing positive or negative feedback, purportedly from the other group, before an intergroup interaction. We expected that perceived interaction quality would predict

outgroup attitudes for participants with positive expectations but not for participants with negative expectations.

Method

Participants and design. One hundred and four participants (64 women) met in same-sex sessions ($n = 20$) of four participants per session in a two-level (expectations: positive, negative) between-subjects design. In Study 2 and all subsequent studies, sample size was based on the effect sizes found in Study 1.

Procedure. Similar to Study 1, two 2-person groups were scheduled for separate locations. During a mass-pretesting session, participants reported their political affiliation. Two "Democrats" and two "Republicans" were scheduled to meet in different laboratory cubicles. Each group of participants was immediately led to a small room without encountering the other group. Participants were informed that they shared a political affiliation (i.e., Republicans or Democrats) with their partner. Participants then worked together on a team-building activity, the Fallout Shelter task (Klein & Willerman, 1979), for 5 min. This task involved a scenario in which participants imagined a fallout shelter without enough supplies for everyone and participants must choose which people will be cast out of the shelter (e.g., an 80-year-old physician).

Upon completion, participants were informed they would soon be working on another task with members of the other political party and that each group would first have an opportunity to evaluate the other group's solution to the Fallout Shelter task. Participants then received a bogus solution purportedly from the other group and evaluated it on four dimensions (1 = *unacceptable* to 8 = *excellent*: Creative, Thoughtful, Effective, Overall Summary). Participants received bogus feedback from the other dyad about their two-person group's solution based on these same dimensions. In the positive expectations condition, participants received positive feedback, whereas in the negative expectation condition, participants received negative feedback. Participants then reported their expectations of the upcoming interaction (six items, three reverse-scored: I expect the interaction in the next room to be [1 = *not at all* to 7 = *very much*]: Cooperative, Tense [R], Friendly, Competitive [R], Quarrelsome [R], Pleasant; $\alpha = .86$) as a manipulation check. To explore the unique effect of expectations beyond any additional influence positive or negative feedback might have on affect, participants also completed the 20-item Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988).

Next, both groups entered another room and together reached consensus on a second problem, the Winter Survival task. Participants worked on this task for 10 min and then reported perceived interaction quality using the same measure as Study 1 (nine items, $\alpha = .88$) and outgroup attitudes using the same adjectives as Study 1 (four items, $\alpha = .76$).

Results

Fifteen participants (eight from the positive condition and seven from the negative condition) were removed from analysis due to knowing one or more participant(s) in the other group prior to the experiment (remaining $N = 89$). We again used HLM using an approach similar to Study 1. As in Study 1, we hypothesized for participants in the positive expectation condition that ratings of perceived interaction quality would predict outgroup attitudes, whereas for those in the negative expectation condition, perceived interaction quality would not predict these attitudes. All models included dummy-coded Political Party at Level 1, and at Level 2, dummy-coded expectation condition and dummy-coded gender. Gender and political party produced no significant main or interaction effects and were therefore removed from models in Study 2. Including these variables did not alter results.

Expectations and quality. We first examined whether positive or negative feedback about the small group solution led participants to expect a positive or negative interaction, respectively, by regressing expected interaction quality on our model. Our manipulation was successful: Participants receiving positive feedback had more positive expectations of the upcoming interaction ($M = 5.39$, $SD = 0.839$) than those receiving negative feedback ($M = 3.83$, $SD = 0.814$; $\gamma = -1.53$, $SE = 0.190$, $p < .001$, 95% CI = [-1.90, -1.16]). This model explained 91% of the Level 2 variance in expected interaction quality.

Next, we examined whether expectation condition had a direct effect on perceived interaction quality by regressing perceived interaction quality on our model. This model explained 17% of the variance in perceived interaction quality. Results indicated, as predicted, that expectation condition predicted perceived interaction quality ($\gamma = 0.464$, $SE = 0.229$, $p = .054$, 95% CI = [0.015, 0.913]), whereby participants in the positive expectations condition perceived the interactions more positively ($M = 5.84$, $SD = 1.01$) than participants in the negative expectations condition ($M = 5.40$, $SD = 0.878$). An additional analysis, with self-reported expected interaction quality as a Level 1 predictor, further supports this notion, with expected interaction quality reliably predicting perceived interaction quality ($\gamma = 0.293$, $SE = 0.099$, $p = .004$, 95% CI = [0.099, 0.487]). Participants with more favorable expectations reported experiencing more positive interactions.

We also examined the effects of the feedback manipulation on affect. We analyzed positive affect and negative affect separately, as recommended for the PANAS, and we also combined them with appropriate reverse-scoring to produce a single measure. The expectation manipulation did not have an effect on overall affect ($\gamma = 0.022$, $SE = 0.056$, $p = .699$, 95% CI = [-0.088, 0.132]) or negative affect ($\gamma = -0.177$, $SE = 0.118$, $p = .147$, 95% CI = [-0.054, 0.408]). For

positive affect, there was a marginal effect ($\gamma = -0.211$, *SE* = 0.118, *p* = .087, 95% CI = [−0.442, 0.020]), but in the opposite direction than expected: Positive affect was somewhat higher in the negative expectations condition (*M* = 2.82, *SD* = 0.604) than in the positive expectations condition (*M* = 2.60, *SD* = 0.606, on a 1-5 scale).

Outgroup attitudes. We then regressed outgroup attitudes on the model described above, with the addition of a grand-mean centered perceived interaction quality as a Level 1 predictor. We again anticipated that perceived interaction quality would predict outgroup attitudes for those participants with positive expectations, whereas perceived interaction quality would be unrelated to outgroup attitudes for those with negative expectations. This model explained 4% of the Level 1 variance and 38% of the Level 2 variance in outgroup attitudes. There was no main effect of expectation condition ($\gamma = 0.184$, *SE* = 0.202, *p* = .387, 95% CI = [-0.580, 0.212]). As predicted, we found a main effect for perceived interaction quality ($\gamma = 0.387$, *SE* = 0.114, *p* = .001, 95% CI = [0.164, 0.610]) on outgroup attitudes. Also as predicted, dummy-coded expectation condition interacted with perceived interaction quality to predict outgroup attitudes ($\gamma = -0.376$, *SE* = 0.208, *p* = .076, 95% CI = [−0.784, 0.032]), replicating Study 1, although the effect was marginally significant (see Figure 2).

Decomposition of this interaction showed effects in the predicted directions, consistent with Study 1. In the positive expectation condition, perceived interaction quality predicted outgroup attitudes ($\gamma = 0.387$, *SE* = 0.112, *t* = 3.41, *p* < .001), whereas in the negative expectation condition, perceived interaction quality was unrelated to outgroup attitudes ($\gamma = 0.012$, *SE* = 0.174, *t* = 0.068, *p* = .946). This indicates that perceived interaction quality predicted outgroup category attitudes for participants entering an interaction with positive expectations, suggesting generalization of contact effects. Furthermore, lack of a relationship between interaction quality and outgroup attitudes for participants with negative expectations suggests no generalization of contact for those participants.

Similar to Study 1, for participants who perceived the interaction to proceed negatively (−1 *SD*), there was no difference between participants within the positive or negative expectation condition ($\gamma = 0.179$, *SE* = 0.255, *t* = 0.704, *p* = .484). This suggests that positive expectations may not lead to worse outgroup attitudes if the interaction goes poorly, because when the interaction went poorly, participants were no more likely to generalize that more negative experience to the outgroup when they had positive—compared with negative—expectations.

Because our feedback manipulation intended to influence intergroup expectations could also influence participants' affect, in a supplementary analysis, we added measures of positive and negative affect as predictors to the model. Our predicted effects occurred independent of affect participants
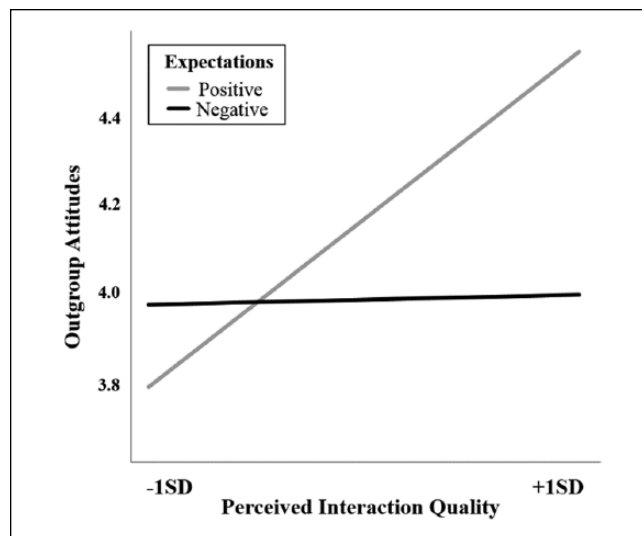

Figure 2. Study 2: Two-way interaction between expectation condition (negative or positive) and perceived interaction quality.

experienced following feedback from the other group. Positive affect did not predict outgroup attitudes ($\gamma = 0.269$, *SE* = 0.214, *p* = .215, 95% CI = [−0.150, 0.688]), but participants who reported more negative affect in response to the other group's feedback subsequently reported more negative outgroup attitudes ($\gamma = -0.765$, *SE* = 0.290, *p* = .011, 95% CI = [−1.33, −0.197]). Of primary relevance to our hypotheses, this analysis yielded the anticipated main effect of perceived interaction quality ($\gamma = 0.343$, *SE* = 0.111, *p* = .003, 95% CI = [0.125, 0.561]) on outgroup attitudes, and the interaction between perceived interaction quality and expectation condition becomes statistically significant ($\gamma = -0.544$, *SE* = 0.190, *p* = .006, 95% CI = [−0.916, −0.172]). These findings indicate that expectations moderate the generalizability of contact above and beyond the influence of affective responses to the feedback manipulation intended to shape expectations.

Discussion

Study 2 provides experimental evidence that positive expectations elicit generalization, whereas negative expectations do not. While participants expecting a more positive interaction rated interactions more positively, these expectations also shape the degree and direction of generalizability to outgroup attitudes. Participants entering an intergroup interaction with positive expectations were open to changing their outgroup evaluation to be consistent with the quality of the interaction, particularly when the interaction proceeded positively. By contrast, perceived interaction quality had no effect on outgroup attitudes for those with negative expectations, no matter how well the interaction itself proceeded.

It is notable that given the contentious political relations between Republicans and Democrats in the United States, experimentally manipulated positive expectations would

make attitudes toward the political outgroup malleable enough to improve with a positive intergroup interaction. Therefore, these results provide support for our hypotheses; nevertheless, preexisting attitudes about the political party outgroups could have affected our results. Likewise, positive and negative expectations were a Level 2 between-subject variable and may have led to qualitative differences in the interactions participants observed. Therefore, in Study 3, we examined the effect of expectations on evaluations of fictitious groups, and we designed the study to eliminate differences in actual interactions that participants in the positive and negative expectation conditions experienced.

Study 3

In Study 3, fictitious groups were utilized to eliminate the influence of any preexisting intergroup biases. To control for objective differences in the interaction, we implemented a yoked design whereby one group in each session would receive the positive expectation manipulation and the other would receive the negative expectation manipulation. So, in each session, participants in both conditions experienced an identical intergroup interaction, but with different expectations. Participant ratings of this interaction thus reflect the effect of the expectation manipulation rather than differences in the interactions when both groups expected positive or negative expectations.

In a two-level (expectations: negative, positive) yoked experimental design, we manipulated expectations in the same way as Study 2, except within the same session we attempted to provide one dyad with positive expectations and the other dyad with negative expectations prior to their interaction. We anticipated replicating Study 2 in that participants expecting a positive interaction would be more sensitive to the interaction quality when evaluating the generalized outgroup compared with those expecting a negative interaction. This hypothesis would be supported by a statistical interaction between expectation condition and perceived interaction quality on outgroup attitudes.

Method

Participants and design. Eighty White undergraduates (48 women) participated for partial course credit, in same-sex sessions ($n = 20$) of four participants per session in a two-level (expectations: positive, negative) between-subjects yoked design.

Procedure. Similar to Study 1, two groups of two participants each were scheduled in separate locations around a central lab. Identical to Study 2, each group of participants was immediately led to a small room without encountering the other group. Participants were informed that based on information collected during a pre-testing session at the beginning the semester, they had been identified as either "Global"

thinkers, those tending to solve problems by seeing the larger picture, or "Analytic" thinkers, those who focus on the smaller details of a problem, and that the purpose of the study was to compare their answers with the other type of thinkers. The two-person groups then worked in separate rooms on the Winter Survival task (see Study 1) for 5 min.

Then, as in Study 2, participants were informed they would soon be working on another task with two outgroup members (Global, Analytic thinkers). Also as in Study 2, participants were informed that before this activity, each group would have an opportunity to evaluate the other group's solution to the Winter Survival task. Valence of evaluative feedback was manipulated in the same way as Study 2. Unlike Study 2, in each session, one group received positive feedback, and the other group received negative feedback. Therefore, in each session, one group expected a positive interaction, whereas the other expected a negative interaction, allowing the actual interactions to be identical for dyads with different expectations. Following receipt of the feedback, participants then reported their expectations (six items, three reverse-scored: I expect the interaction in the next room to be [1 = *not at all* to 7 = *very much*]: Cooperative, Tense [R], Friendly, Competitive [R], Quarrelsome [R], Pleasant; α = .83), as a manipulation check.

Next, both groups then entered another room and worked together on a Fallout Shelter task to foster group discussion (see Study 2). Participants worked on this task for 10 min, following which they completed the same measure of perceived interaction quality utilized in Study 1 (nine items, α = .87) and a measure of perceptions of Global and Analytical thinkers in general (four items; α = .76).

Results

Again using HLM, for all models we entered dummy-coded expectation condition (expectations: negative, positive) at Level 1 and gender at Level 2. Unless explicitly stated, gender produced no significant effects and was therefore removed from models in Study 2. The inclusion of gender did not alter results. As a manipulation check, we examined whether positive or negative feedback about the small group solution led participants to expect a positive or negative interaction, respectively. Regressing expected interaction quality on our model indicated that our manipulation was successful: Participants receiving positive feedback had more positive expectations of the upcoming interaction ($M = 5.59$, $SD = 0.710$) than those receiving negative feedback ($M = 3.93$, $SD = 0.873$; $\gamma = 1.66$, $SE = 0.186$, $p < .001$, 95% CI = [1.30, 2.02]). This model explained 57% of the Level 1 variance in expected interaction quality.

Next, we regressed perceived interaction quality on our model. As predicted and confirming that our yoked design indeed controlled for a direct effect of expectations on interaction quality, there was no effect of expectation condition on perceived interaction quality ($\gamma = 0.228$, $SE = 0.160$, $p =$
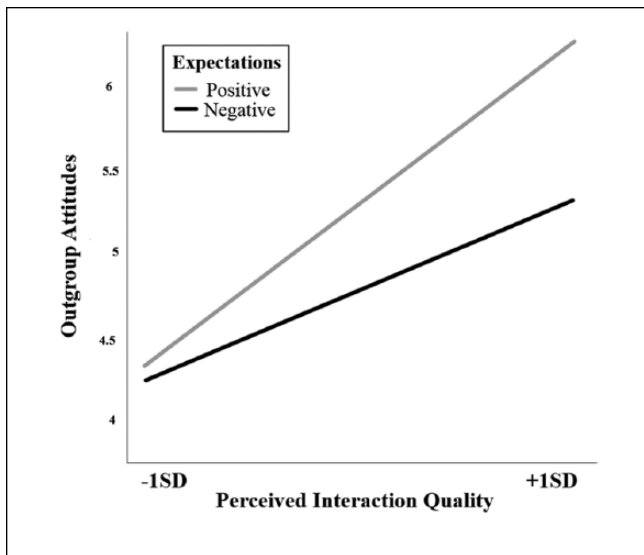
Figure 3. Study 3: Two-way interaction between expectation condition (negative or positive) and perceived interaction quality.

.123, 95% CI = [−0.086, 0.542]). Perceived interaction quality was generally positive across all conditions ($M = 6.01$, $SD = 0.759$). We also tested a model including group membership to determine whether both groups had a similar perception of the interaction. The results indicated that there was no main effect for group membership (Analytic, Global thinkers) or Group membership × Expectation interaction in perceived interaction quality.

Outgroup attitudes. We predicted that positive expectations would encourage generalization, whereas negative expectations would inhibit it. Thus, like in Studies 1 and 2, we hypothesized that ratings of perceived interaction quality would interact with expectation condition to predict outgroup attitudes. Therefore, we regressed outgroup category attitudes on our statistical model, including grand-mean centered perceived interaction quality and the Expectation condition × Perceived interaction quality multiplicative interaction term as Level 1 predictors. This model explained 44% of the Level 1 variance in outgroup attitudes. We found a main effect for expectation condition ($\gamma = 0.486$, $SE = 0.192$, $p = .014$, 95% CI = [0.110, 0.862]) and a main effect for perceived interaction quality ($\gamma = 0.684$, $SE = 0.204$, $p = .001$, 95% CI = [0.284, 1.08]). As predicted, expectation condition interacted with perceived interaction quality to predict outgroup attitudes ($\gamma = 0.553$, $SE = 0.219$, $p = .014$, 95% CI = [0.101, 0.959]), replicating Studies 1 and 2 (see Figure 3). The simple slopes showed that perceived interaction quality predicted outgroup attitudes in the positive expectation condition ($\gamma = 1.24$, $SE = 0.126$, $t = 9.82$, $p < .001$). Consistent with our hypothesis, this relationship was of lesser magnitude in the negative expectation condition ($\gamma = 0.684$, $SE = 0.204$, $t = 3.35$, $p = .002$).

As with Studies 1 and 2, when perceived interaction quality was low (−1 *SD*), there was no difference between the negative expectations and the positive expectations conditions ($\gamma = 0.067$, $SE = 0.241$, $t = 0.279$, $p = .781$). Therefore, our hypothesis that positive expectations would lead to greater generalization of negative contact, in addition to positive contact, was not supported. This pattern again suggests that positive expectations encourage generalization of positive contact to improve intergroup attitudes but generalize no differently than negative expectations when the contact situation is less amiable.

Discussion

Study 3 provides further support for our hypotheses and replicates Studies 1 and 2. Although both negative and positive expectations showed a positive relationship between perceived interaction quality and outgroup attitudes, this relationship was qualified by a significant interaction between expectation condition and perceived interaction quality. Consistent with Studies 1 and 2, this statistical interaction suggests that positive expectations encouraged generalization. The relationship between perceived interaction quality and outgroup category attitudes was more pronounced in the positive expectation condition, indicating that participants with positive expectations more readily associated outcomes from that interaction with the outgroup, whereas negative expectations did not.

Given that participants expecting both positive and negative interactions experienced the same group discussion, the effect of expectations on generalization was likely due to differences in processing rather than differences in the actual interaction. This result is particularly noteworthy given the use of fictitious groups to which participants had no previous association. Participants interacting with members of established outgroups (racial outgroup members in Study 1 and political rivals in Study 2) are likely more motivated to retain negative outgroup category attitudes. The use of fictitious groups in this study both eliminates the effect of preexisting attitudes on evaluations of the outgroup and highlights the importance of expectations in generalizing a positive contact experience to the outgroup.

The strength of our yoked design is that, regardless of expectations, participants experienced the same actual interaction, providing greater experimental control. As predicted, ratings of interaction quality were the same for those with positive and negative expectations. However, although participants rated it equivalently in both conditions, we had no control over interaction quality. Although expectations had an effect on generalization of contact, it did not have a direct effect on perceptions of the interaction quality. This may have been due to the yoked experimental design, which was intended to minimize interaction differences between participants with positive or negative expectations for the interaction. Without manipulating the interaction quality directly,

our design may have restricted the range of perceived interaction quality ratings. A generally positive perceived interaction quality—in this study, $M = 6.01$ out of 7—may have prevented us from testing the effect of a truly negative intergroup interaction and may have washed out any potential effect of expectations on perceived interaction quality. Therefore, we designed Study 4 to manipulate both the objective interaction quality and participants' expectations.

Study 4

In Study 4, we manipulated both the expectations for the interaction and the quality of the *actual* interaction using trained undergraduate confederates. This design enabled us to directly test the interactive effect of participant expectations and the actual interaction quality on outgroup perceptions.

Method

Participants and design. Fifty-eight White undergraduate women[2] participated for partial course credit, in 29 sessions of two participants and two confederates per session, in a 2 (expectations: positive, negative) × 2 (interaction quality: positive, negative) between-subjects design.

Confederate behavior. To manipulate interaction quality, each session included two confederates posing as participants. In the positive interaction condition, the confederates acted to facilitate an agreeable intergroup interaction, whereas in the negative interaction condition, confederates acted to make it difficult to reach a final decision. Confederates were trained to use prepared scripts of six statements, three from each confederate. In the *positive interaction condition*, these statements were variations of the following: "That is true, [item] is definitely the most important," "I agree, that is a good spot for [item]," Great idea, I didn't think of that," and "That is pretty creative. I like that." In the *negative interaction condition*, these statements comprised two direct criticisms of a suggestion, and four suggestions to reorganize the rank ordering of the list after items had already been recorded, intended to be irritating, such as "Actually I think [item] should go as #1. Let's just move everything down one rank," "Now that I think about it, I don't think that's a good choice for #1. What if we put [item] there?," "Actually, [item] doesn't make sense at #4, let's move it to #2," and "I don't think that's a good idea, but whatever, we can just keep going." Other than variations of these statements, confederates were trained to keep their comments to a minimum. Confederates were unaware of the expectation condition.

Procedure. Two confederates waited in the same location as participants before an experimenter ushered them into the laboratory. To increase believability of the fictitious

under- and over-estimator groups, each of the four ostensible participants was asked to estimate of the number of jellybeans of a particular color in a jar of jellybeans. Both actual participants were informed they belonged to one group (over- or under-estimators) and both confederates belonged to the other group. The two participants were then led to a small room, where they were informed the purpose of the experiment was to compare the problem-solving styles of over- and under-estimators. Participants then worked on the Fallout Shelter task for 5 min.

Similar to Study 2 and Study 3, expectations were manipulated by participants either receiving positive or negative feedback on their solution, and participants then reported their expectations of the upcoming interaction (expected interaction quality, 11 items, six reverse-scored: Cooperative, Friendly, Agreeable, Successful, Decisive, Tense [R], Competitive [R], Quarrelsome [R], Unpleasant [R], Frustrating [R], Annoying [R]; α = .89).

Similar to Studies 1 and 2, participants then entered another room with the two confederates, and each group worked together on the Winter Survival task for 10 min. Participants were told the four-person group creating the best solution over the semester would receive US$40.00. To manipulate interaction valence, the confederates acted in the manner described earlier. Following the interaction, they completed a questionnaire asking about perceived interaction (11 items, same as expected interaction quality, α = .95) and of the outgroup in general (either over- or under-estimators; 4 items, α = .82).

Results

Two participants were eliminated due to suspicion of the feedback (remaining $N = 56$). Unlike Studies 1 to 3, in the current study, expectations of the upcoming interaction and interaction quality were manipulated between subjects. Therefore, there were no Level 1 predictors, and grand-mean centered Level 2 predictors included dummy-coded expectation condition (positive, negative) and dummy-coded interaction quality condition (positive, negative) and their multiplicative interaction term.

As a manipulation check, we regressed expected interaction quality on our model. Participants in the negative expectations conditions ($M = 5.36$, $SD = 0.863$) expected a worse interaction than participants in the positive expectations conditions ($M = 5.36$, $SD = 0.863$; $\gamma = 1.05$, $SE = 0.312$, $p = .002$, 95% CI = [0.438, 1.66]). This model explained 69% of the Level 2 variance in expected interaction quality. Next, as a manipulation check, we regressed perceived interaction quality on our model. This model explained 88% of the Level 2 variance in perceived interaction quality. Participants in the negative interaction conditions ($M = 4.20$, $SD = 1.05$) evaluated the interaction more negatively than did those in the positive interaction conditions ($M = 6.45$, $SD = 0.473$; $\gamma = 2.06$, $SE = 0.237$, $p < .001$, 95% CI = [1.60, 2.52]), indicating
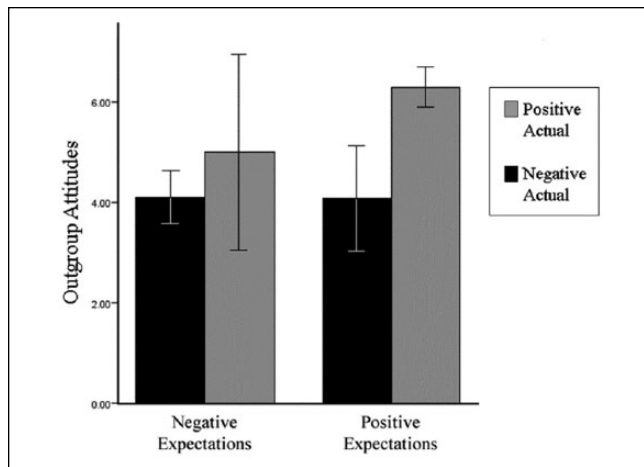
Figure 4. Expectation (negative or positive) by actual interaction (negative or positive) interaction.
Note. Error bars represent 95% confidence intervals.

that both manipulations were successful. Moreover, participants in the negative expectations conditions did not differ from participants in the positive expectations conditions ($\gamma = -0.039$, $SE = 0.375$, $p = .918$, 95% CI = [−0.774, 0.696]) on ratings of perceived interaction quality, indicating that participant expectations did not have a direct effect on perceived interaction quality.[3]

Outgroup attitudes. To examine the effect of expected and perceived interaction quality on outgroup attitudes, we regressed outgroup category attitudes on our model. This model explained 68% of the Level 2 variance in outgroup category attitudes. We found neither an effect of expectation condition ($\gamma = -0.338$, $SE = 0.406$, $p = .413$, 95% CI = [−1.13, 0.458]) nor an effect of interaction condition ($\gamma = 0.743$, $SE = 0.492$, $p = .143$, 95% CI = [−0.221, 1.71]) on outgroup attitudes. Consistent with the previous studies, we did find the anticipated Expectation condition × Interaction condition statistical interaction on outgroup attitudes ($\gamma = 1.41$, $SE = 0.664$, $p = .043$, 95% CI = [0.109, 2.71]). As predicted and replicating Studies 1, 2, and 3, decomposing the interaction revealed that in the negative expectations condition, perceived interaction quality did not influence evaluations of the outgroup ($\gamma = 0.743$, $SE = 0.492$, $t = 1.51$, $p = .143$), and in the positive expectations condition, perceived interaction quality predicted outgroup attitudes ($\gamma = 2.16$, $SE = 0.446$, $t = 4.84$, $p < .001$; see Figure 4).

Consistent with the previous studies, when the intergroup interaction was negative, outgroup attitudes for participants with negative expectations were not significantly different from that for participants with positive expectations ($\gamma = -0.034$, $SE = 0.406$, $t = -0.832$, $p = .414$). In Study 4, expectations did not have an effect on perceived interaction quality, a result that is likely due to our experimental manipulation of interaction quality.

Discussion

Study 4 provides strong evidence for our hypothesis that positive expectations facilitate, whereas negative expectations limit generalization of an intergroup interaction on outgroup attitudes. In Study 4, expectations did not have an effect on perceived interaction quality, a result that is likely due to our direct manipulation of interaction quality. For participants in the positive expectation condition, those who had a positive interaction showed more positive outgroup attitudes compared with those who had a negative interaction. In contrast, for participants in the negative expectations condition, outgroup attitudes did not differ as a function of how well the interaction actually proceeded. Study 4 sampled only female participants, but because our results were consistent with the previous three studies and given that gender did not play a role in the other studies, we do not believe this to be a factor in our results.

Importantly, Study 4 replicated the effects observed in each of the previous three studies. First, we showed that when people have positive expectations—but not when they have negative expectations—positive interactions lead to more positive outgroup attitudes compared with participants who experienced negative interactions, suggesting that positive expectations encourage generalization. Given that we manipulated both expectations and interaction quality, this study provides evidence that having positive expectations for an interaction causes outgroup attitudes to be dependent on how well that interaction proceeds. For participants expecting a positive interaction, a positive intergroup interaction can produce more favorable outgroup views. By contrast, having negative expectations for an intergroup interaction may crystallize outgroup attitudes, making them less responsive to a positive intergroup interaction.

General Discussion

Using correlational, yoked quasi-experimental, and experimental designs involving both minimal and real-world groups, four studies demonstrated that expectations about intergroup interactions, whether measured (Study 1) or manipulated (Study 2), predicted the interaction quality. Only in studies where the design of our study necessarily limited (yoked design, Study 3) or directly manipulated (Study 4) the interaction quality did we fail to find a main effect of expectations on interaction quality. We believe the critical difference in the findings between these two sets of studies is that in Studies 1 and 2, we allowed the interaction to vary spontaneously, whereas in Studies 3 and 4, we more tightly controlled the valence of the interactions. Interactions with no direct valence manipulation (Studies 1 and 2) are more likely to reflect naturalistic interactions beyond the laboratory, and thus we argue that these studies have greater external validity, though less empirical precision. Thus, expectations usually influence perceived interaction quality.

The unique contribution of the current research is the role that expectations play in generalizing intergroup contact to outgroup attitudes. Factors such as the typicality of outgroup members directly involved in the contact situation has been shown to moderate generalization of positive contact to improved outgroup attitudes (see Brown & Hewstone, 2005, for a review). The current research demonstrates that intergroup expectations also moderate the extent to which the quality of intergroup contact affects outgroup attitudes. Whereas positive intergroup expectations encourage generalization of positive contact to the outgroup category, negative expectations help maintain negative outgroup attitudes. Across all four studies, participants with positive expectations for the intergroup interaction generalized a positive intergroup interaction to the outgroup category. By contrast, for participants holding negative expectations, the relationship between intergroup interaction quality and outgroup attitudes was nonexistent or weaker than for participants with positive expectations.

These studies not only highlight the importance of positive intergroup contact but also reinforce recent interest in negative contact. A recent longitudinal study suggests that both positive and negative intergroup contact can generalize to the outgroup (Stark, Flache, & Veenstra, 2013). Additional research suggests that negative information may more easily generalize to the outgroup (Paolini, Harwood, & Rubin, 2010). Our studies suggest that negative expectations prevent positive contact from leading to more favorable outgroup attitudes.

One unexpected finding was consistent. Positive expectations did not appear to have a drawback in negative contact situations, compared with negative expectations. Although we predicted that positive expectations might make people more likely to generalize a negative contact experience, in all studies, when the actual interaction was relatively negative, there was no difference between participants with positive or negative expectations. Future studies should use a baseline control condition to test whether positive or negative expectations specifically cause changes in the likelihood of generalization. For example, in the paradigm used in Studies 2 and 3, a condition could be added in which the groups receive no information about what the other group thought of their group's solution.

Positive expectations appear to be effective at improving intergroup relations (Mallett & Wilson, 2010). However, affective forecasting research suggests that people are generally poor predictors of their feelings (Wilson & Gilbert, 2003), and people systematically predict that intergroup interactions will proceed worse than they actually do (Mallett et al., 2008). People tend to adjust predictions based on negative feedback, but they tend not to learn from positive feedback (Wilson, Meyers, & Gilbert, 2001). This may explain why negative expectations are so impactful and why they tend to stifle generalization in the face of positive contact. Moreover, this may suggest that people will fail to adjust

their expectations of future contact following a positive contact situation, particularly for people who entered an interaction with negative expectations.

Limitations and Future Directions

One potential limitation to our research is the use of small groups in each study, instead of individual participants. Research on the interindividual-intergroup discontinuity effect suggests that people behave more competitively when in groups compared with when interacting as individuals (for a review, see Wildschut, Pinter, Vevea, Insko, & Schopler, 2003). However, because our studies were intended to test the effect of expectations for members of different social group categories, we would suspect on the basis of social identity theory (Tajfel & Turner, 1979) that the results would be no different than if participants were interacting as two individuals, as long as their different social identities were salient. When individuals' group identities are salient, two individuals can have an intergroup interaction (Brown & Turner, 1985). However, future research might examine potential differences in the role expectations play when individuals with different salient group identities, compared with small groups, interact.

Although the present work provides convergent evidence across four studies of the role of intergroup expectations on the way the quality of intergroup contact affect outgroup attitudes, we do not examine a mediator of the effect, which future research might consider. Information-processing styles are a possible mediator. Positive expectations facilitate greater scrutiny toward an individual than negative expectations (Yzerbyt & Leyens, 1991), which entrench negative attitudes toward that individual (Ybarra, 2002), making attitude change less likely. We observed this pattern in our measures capturing the effect of expectations and the interaction quality on outgroup attitudes. That is, positive intergroup expectations led to greater generalization from the interaction to outgroup attitudes, compared with negative expectations. Positive expectations elicit a more systematic processing style, whereas negative expectations produce a more heuristic processing style (Ybarra et al., 1999). It is possible that intergroup expectations have an effect on intergroup information-processing styles, leading to differences in how easily a positive contact experience generalizes to the outgroup (Ybarra, Stephan, & Schaberg, 2000).

Additional research might directly assess whether positive and negative expectations lead people to process intergroup interactions with deliberative or heuristic processing. To the extent that deliberative processing accounts for the effects of positive expectations on the generalization of the immediate intergroup experience to outgroup attitudes, introducing memory load on participants, which reduces the capacity for deliberative processing, might impair the effects of positive expectations on generalization but have little effect on the pattern we observed for negative expectations.

Deliberative processing could be assessed, as Ybarra et al. (1999) did, by administering a memory test after presenting the information about social targets. Specifically, future research could script confederate behavior to convey identical information across conditions. Better memory for this information is assumed to reflect more deliberative processing. We would expect that positive, compared with negative, intergroup expectations would produce better memory for the detailed information conveyed by confederates, representing greater engagement and scrutiny during an interaction, which in turn would predict a stronger relationship between the specific attitudes toward outgroup members involved in the interaction and the outgroup category. Directly assessing the potential underlying cognitive processes would thus help to clarify the relationship between intergroup expectations and post-interaction intergroup attitudes.

Another limitation of this research is the use of a feedback-based manipulation of expectations. Although our manipulation checks in Studies 2, 3, and 4 indicated that we successfully manipulated perceived expectations, the possibility remains that other potential processes could have been manipulated. In Study 2, our results were due to changes in perceived expectations above and beyond the effect of affect, though we have no data available in Studies 3 and 4 to examine the potential role of affect. The use of just one method of manipulating expectations represents a limitation of this research, and future studies should use other manipulations and study designs to more directly examine the role of expectations in generalization of intergroup contact.

In addition, future research might examine the effect of expectations on performance during an intergroup interaction. Positive performance expectancies generally lead to better performance (Bandura, 1986), and therefore, one might predict that positive expectations might encourage better performance. However, other research suggests that expectations need to be realistically positive (i.e., not fantasies) to be effective (Oettingen & Mayer, 2002) and that task difficulty and cognitive capacity (Reinhard & Dickhäuser, 2009) must be taken into account for positive expectations to lead to improved performance. Moreover, the challenges posed by intergroup interactions certainly play a role, suggesting that the relationship between expectations and performance in an intergroup context may be more nuanced.

Across these four studies, participants experienced a generally positive interaction. Because the interactions were generally positive, which would be anticipated based on the cooperative problem-solving tasks used, participants with positive expectations likely had their expectations confirmed more than those with negative expectations. Consequently, they may feel more confident in their intergroup interactions, perhaps reducing intergroup anxiety, which negatively influences outgroup attitudes (Plant & Devine, 2003). Future research might consider feelings of confidence with intergroup interactions, as well as intergroup anxiety, as potential mediators of the expectations by interaction quality effects on outgroup attitudes.[4] In addition, future research might more strongly manipulate the favorability of the intergroup interaction. If correspondence between expectations and the experience of the intergroup interaction is crucial, then positive expectations would be expected to primarily promote generalization when the experience of the intergroup interaction was positive, but negative expectations would primarily produce generalization when the intergroup interaction was negative.

In conclusion, although positive intergroup contact plays a crucial role in encouraging group-level attitude change (Pettigrew et al., 2011), positive contact does not necessarily guarantee more positive intergroup attitudes. Expectations fundamentally shape the ways people attend to their environment, encode information into memory, and form attitudes—particularly about other people. Unsurprisingly, positive expectations appear to have a very fruitful role in improving intergroup relations (Mallett & Wilson, 2010). The current research indicates that while more positive expectations have a positive impact on intergroup interactions, they also encourage greater generalization of an interaction to the outgroup. This research also reveals the particularly damaging role of negative expectations, which prevents generalization. Thus, the content of one's expectations upon entering an intergroup interaction can either facilitate or limit the potentially beneficial effects of positive contact on intergroup attitudes.

Declaration of Conflicting Interests

Funding

Notes

1. Currently, no consensus exists on which effect size measures to use in multilevel analyses (Peugh, 2010). A local effect size can be calculated for Level 1 and Level 2 variance components representing the proportion of variance reduction from a model including no predictors. The proportion of variance reduction allows us to show the amount of variance in the outcome variable that is explained by the entire model at either Level 1 or Level 2, depending on whether variables are entered at each level (Peugh, 2010; Raudenbush & Bryk, 2002). Therefore, for all models reported, we have included this statistic.
2. To maintain consistency between sessions, four individuals acted as confederates. These confederates were all female, and therefore, only female participants were recruited.
3. An additional analysis examining the effect of self-reported expectations produced similar results.

4. We conducted additional analyses to test, albeit indirectly, whether greater confidence accounts for some of our findings. We computed an accuracy score using the difference (in absolute value) between expected interaction quality and perceived interaction quality (in Studies 2 and 3, which included different items for expectations and interaction quality, we used the items that were common to both measures; no adjustment was needed for Studies 1 and 4). In these analyses, greater correspondence between expectations and perceived interaction quality (or expectation "accuracy") is assumed to elicit greater confidence about interacting with outgroup members. We tested the relationship between accuracy (low scores indicating less difference or greater accuracy) and outgroup attitudes within our nested design for each study. We found inconsistent support for an interpretation that greater confidence may mediate our results. Greater accuracy predicted more positive outgroup attitudes in Study 1 ($\gamma = -0.487$, $SE = 0.172$, $p = .007$, 95% CI = $[-0.824, -0.150]$) and in Study 2 ($\gamma = -0.322$, $SE = 0.147$, $p = .032$, 95% CI = $[-0.610, -0.034]$). The relationship was nonsignificant in Study 3 ($\gamma = -0.265$, $SE = 0.174$, $p = .133$, 95% CI = $[-0.606, 0.076]$). In Study 4, by contrast, the lower the difference between expected and actual interaction quality, the more *negatively* participants rated the outgroup ($\gamma = 0.446$, $SE = 0.128$, $p = .002$, 95% CI = $[0.195, 0.697]$).

Supplemental Material

The online supplemental material is available at http://pspb.sagepub.com/supplemental.

References

Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, 323-370.

Brown, R. J., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in Experimental Social Psychology*, 37, 255-343.

Brown, R. J., & Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behavior. *Advances in Group Processes*, 2, 33-65.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62-68.

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology*, 57, 239-249.

Gotlib, I. H., Kasch, K. L., Traill, S., Joormann, J., Arnow, B. A., & Johnson, S. L. (2004). Coherence and specificity of information-processing biases in depression and social phobia. *Journal of Abnormal Psychology*, 113, 386-398.

Johnson, D. W., & Johnson, E. W. (1975). *Joining together: Group theory and group skills*. Englewood Cliffs, NJ: Prentice Hall.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. *Advances in Experimental Social Psychology*, 2, 219-266.

Klein, H. M., & Willerman, L. (1979). Psychological masculinity and femininity and typical and maximal dominance expression in women. *Journal of Personality and Social Psychology*, 37, 2059-2070.

Lingle, J. H., & Ostrom, T. M. (1979). Retrieval selectivity in memory-based impression judgments. *Journal of Personality and Social Psychology*, 37, 180-194.

Maass, A., Milesi, A., Zabbini, S., & Stahlberg, D. (1995). Linguistic intergroup bias: Differential expectancies or in-group protection? *Journal of Personality and Social Psychology*, 68, 116-126.

Mallett, R. K., & Wilson, T. D. (2010). Increasing positive intergroup contact. *Journal of Experimental Social Psychology*, 46, 382-387.

Mallett, R. K., Wilson, T. D., & Gilbert, D. T. (2008). Expect the unexpected: Failure to anticipate similarities leads to an intergroup forecasting error. *Journal of Personality and Social Psychology*, 94, 265-277.

Oettingen, G., & Mayer, D. (2002). The motivating function of thinking about the future: Expectations versus fantasies. *Journal of Personality and Social Psychology*, 83, 1198-1212.

Paolini, S., Harwood, J., & Rubin, M. (2010). Negative intergroup contact makes group memberships salient: Explaining why intergroup conflict endures. *Personality and Social Psychology Bulletin*, 36, 1723-1738.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49, 65-85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90, 751-783.

Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations*, 35, 271-280.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48, 85-112.

Plant, E. A., Butz, D. A., & Tartakovsky, M. (2008). Interethnic interactions: Expectancies, emotions, and behavioral intentions. *Group Processes & Intergroup Relations*, 11, 555-574.

Plant, E. A., & Devine, P. G. (2003). The antecedents and implications of interracial anxiety. *Personality and Social Psychology Bulletin*, 29, 790-801.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: SAGE.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2008). *HLM: Hierarchical linear and nonlinear modeling* (Version 6.06). Lincolnwood, IL: Scientific Software International.

Reinhard, M. A., & Dickhäuser, O. (2009). Need for cognition, task difficulty, and the formation of performance expectancies. *Journal of Personality and Social Psychology*, 96, 1062-1076.

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review*, 10, 336-353.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues*, 41(3), 81-104.

Shelton, J. N., & Richeson, J. A. (2005). Intergroup contact and pluralistic ignorance. *Journal of Personality and Social Psychology*, 88, 91-107.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behaviour: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, *35*, 656-666.

Stark, T. H., Flache, A., & Veenstra, R. (2013). Generalization of positive and negative attitudes toward individuals to outgroup attitudes. *Personality and Social Psychology Bulletin*, *39*, 608-622.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063-1070.

West, T. V., & Dovidio, J. F. (2012). Intergroup contact across time: Beyond initial contact. In G. Hodson & M. Hewstone (Eds.), *Advances in intergroup contact* (pp. 152-175). New York, NY: Psychology Press.

Wildschut, T., Pinter, B., Vevea, J. L., Insko, C. A., & Schopler, J. (2003). Beyond the group mind: A quantitative review of the interindividual-intergroup discontinuity effect. *Psychological Bulletin*, *129*, 698-722.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. *Advances in Experimental Social Psychology*, *35*, 345-411.

Wilson, T. D., Meyers, J., & Gilbert, D. T. (2001). Lessons from the past: Do people learn from experience that emotional reactions are short-lived? *Personality and Social Psychology Bulletin*, *27*, 1648-1661.

Ybarra, O. (2002). Naive causal understanding of valenced behaviors and its implications for social information processing. *Psychological Bulletin*, *128*, 421-441.

Ybarra, O., Schaberg, L., & Keiper, S. (1999). Favorable and unfavorable target expectancies and social information processing. *Journal of Personality and Social Psychology*, *77*, 698-709.

Ybarra, O., Stephan, W. G., & Schaberg, L. (2000). Misanthropic memory for the behavior of group members. *Personality and Social Psychology Bulletin*, *26*, 1515-1525.

Yzerbyt, V. Y., & Leyens, J. P. (1991). Requesting information to form an impression: The influence of valence and confirmatory status. *Journal of Experimental Social Psychology*, *27*, 337-356.

Universal dimensions of social cognition: warmth and competence

Susan T. Fiske[1], Amy J.C. Cuddy[2] and Peter Glick[3]

[1] Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08540, USA
[2] Management and Organizations Department, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, USA
[3] Psychology Department, Lawrence University, PO Box 599, Appleton, WI 54912, USA

Like all perception, social perception reflects evolutionary pressures. In encounters with conspecifics, social animals must determine, immediately, whether the 'other' is friend or foe (i.e. intends good or ill) and, then, whether the 'other' has the ability to enact those intentions. New data confirm these two universal dimensions of social cognition: warmth and competence. Promoting survival, these dimensions provide fundamental social structural answers about competition and status. People perceived as warm and competent elicit uniformly positive emotions and behavior, whereas those perceived as lacking warmth and competence elicit uniform negativity. People classified as high on one dimension and low on the other elicit predictable, ambivalent affective and behavioral reactions. These universal dimensions explain both interpersonal and intergroup social cognition.

Introduction

Dark alleys and battle zones approximate the survival settings of ancestral encounters with strangers. Evolutionary pressures are reflected in social perception: on encountering others, people must determine, first, the intentions of the other person or group and, second, their ability to act on those intentions. In the past few years, research has clearly established that perceived warmth and competence are the two universal dimensions of human social cognition, both at the individual level and at the group level. The evidence for these dimensions comes from various sources, including experimental social psychology laboratories, election polls and cross-cultural comparisons. Decades of prior research supports the importance (and constant recurrence) of the warmth and competence dimensions, under various labels (Box 1). However, only in the past five years have cutting-edge studies of social cognition firmly established that people everywhere differentiate each other by liking (warmth, trustworthiness) and by respecting (competence, efficiency).

According to recent theory and research in social cognition, the warmth dimension captures traits that are related to perceived intent, including friendliness, helpfulness, sincerity, trustworthiness and morality, whereas the competence dimension reflects traits that are related to perceived ability, including intelligence, skill, creativity and efficacy. For example, these dimensions appear in

spontaneous impressions of presidential candidates, which entail both competence and integrity (warmth, trustworthiness) [1–3]. Impressions of leaders also involve these dimensions and include image management (building trust), relationship development (warmth) and resource deployment (competence and efficacy) [4]; although one could quibble over separating or combining trust and warmth, there is a core linkage between the two features, with trust and warmth consistently appearing together in the social domain.

These public-sector results are borne out by studies from Bogdan Wojciszke's laboratory on how people construe the behavior of others. The basic dimensions of warmth and competence account for 82% of the variance in perceptions of everyday social behaviors [5]. Three-quarters of more than 1000 personally experienced past events are framed in terms of either morality or competence [6], and impressions of well-known people show a similar pattern [5] (reviewed in Ref. [7]). The terms used by Wojciszke and colleagues [5,6] are translated as 'competence' and 'morality', but the moral traits include fair, generous, helpful, honest, righteous, sincere, tolerant and understanding, which overlap with the warmth–trustworthiness dimension that has been identified elsewhere. (There is no dispute over the competence label; these traits include clever, competent, creative, efficient, foresighted, ingenious, intelligent and knowledgeable.) In sum, when people spontaneously interpret behavior or form impressions of others, warmth and competence form basic dimensions that, together, account almost entirely for how people characterize others.

The primacy of warmth judgments

Although warmth and competence dimensions emerge consistently, considerable evidence suggests that warmth judgments are primary: warmth is judged before competence, and warmth judgments carry more weight in affective and behavioral reactions. From an evolutionary perspective, the primacy of warmth is fitting because another person's intent for good or ill is more important to survival than whether the other person can act on those intentions. Similarly, morality (warmth) judgments determine approach–avoidance tendencies, so they are the fundamental aspect of evaluation [8,9] and, therefore, precede competence–efficacy judgments. People infer warmth from the perceived motives of the other person [10]. Information

Corresponding author: Fiske, S.T. (sfiske@princeton.edu).
Available online 22 December 2006.

Box 1. History of research on person perception

In 1946, Solomon Asch [62] published a paradigmatic study in which undergraduates formed impressions of another person based on lists of trait adjectives (e.g. determined, practical, industrious, intelligent, skillful), which also included either 'warm' or 'cold' depending on the experimental condition. The power of warm versus cold as 'central traits' that dramatically alter impressions has been the stuff of introductory textbooks ever since. These impression studies demonstrated the role of Gestalt clusters in social perception: a warm intelligent person is wise, whereas a cold intelligent person is sly.

Decades later, assuming that certain traits tend to separate into clusters, Rosenberg *et al.* [22] asked undergraduates to sort 64 traits into categories that are likely to be associated in the same person. Multidimensional scaling and subsequent analyses identified two primary dimensions: social good–bad and intellectual good–bad. As Figure I indicates, socially good traits include warm (as found in Ref. [62]) and sociable (as found in Ref. [22]), plus good-natured, happy, popular and sincere; socially bad traits are their opposites. Nearly orthogonal are the intellectually good–bad traits: intelligent, scientific, persistent, determined, skillful and industrious, and their opposites. Asch's dramatic results for warm–cold could be explained by the sociability dimension (warm–cold) being varied while the intellectual dimension was kept constant [23]. The warm–cold manipulation gains its power to alter the Gestalt of an impression by tapping a fundamental aspect of how people are perceived.

Nevertheless, the implications of these basic dimensions of person perception did not reach total consensus immediately. Furthermore,

calling trait lists 'person perception' was empirically tractable but ecologically problematic. Some studies (e.g. Ref. [21]) addressed ecological validity by providing pictures of stimulus persons engaged in personality-revealing behaviors on two cognate dimensions, such as sociability and responsibility. However, these laboratory studies entailed experimenter-chosen traits, which capitalized on the apparent distinction between the two dimensions but brought into question the perceivers' spontaneously used dimensions.

Fortunately, in parallel, impressions of others within small, interactive groups were found to include separate social (warmth) and task (competence) orientations [63]. Generations of Harvard university undergraduates in Robert Freed Bales's self-observational small group class and interacting small groups in a variety of organizations converged on these two dimensions [64]. The Bales system included a third dimension – sheer volume of interaction. This is probably most salient in the live interaction context but less salient in stored impressions.

In sum, there is a venerable history of warmth and competence dimensions that emerge in independent lines of research. One could add self-perception to this list (e.g. independent, agentic versus interdependent, communal) in addition to work on perceptions of social categories (e.g. the distinction between communion and agency in gender stereotypes). However, the various labels that have been used for these basic dimensions had (until recently) obscured the pervasiveness and power of the fundamental, underlying dimensions of warmth and competence.
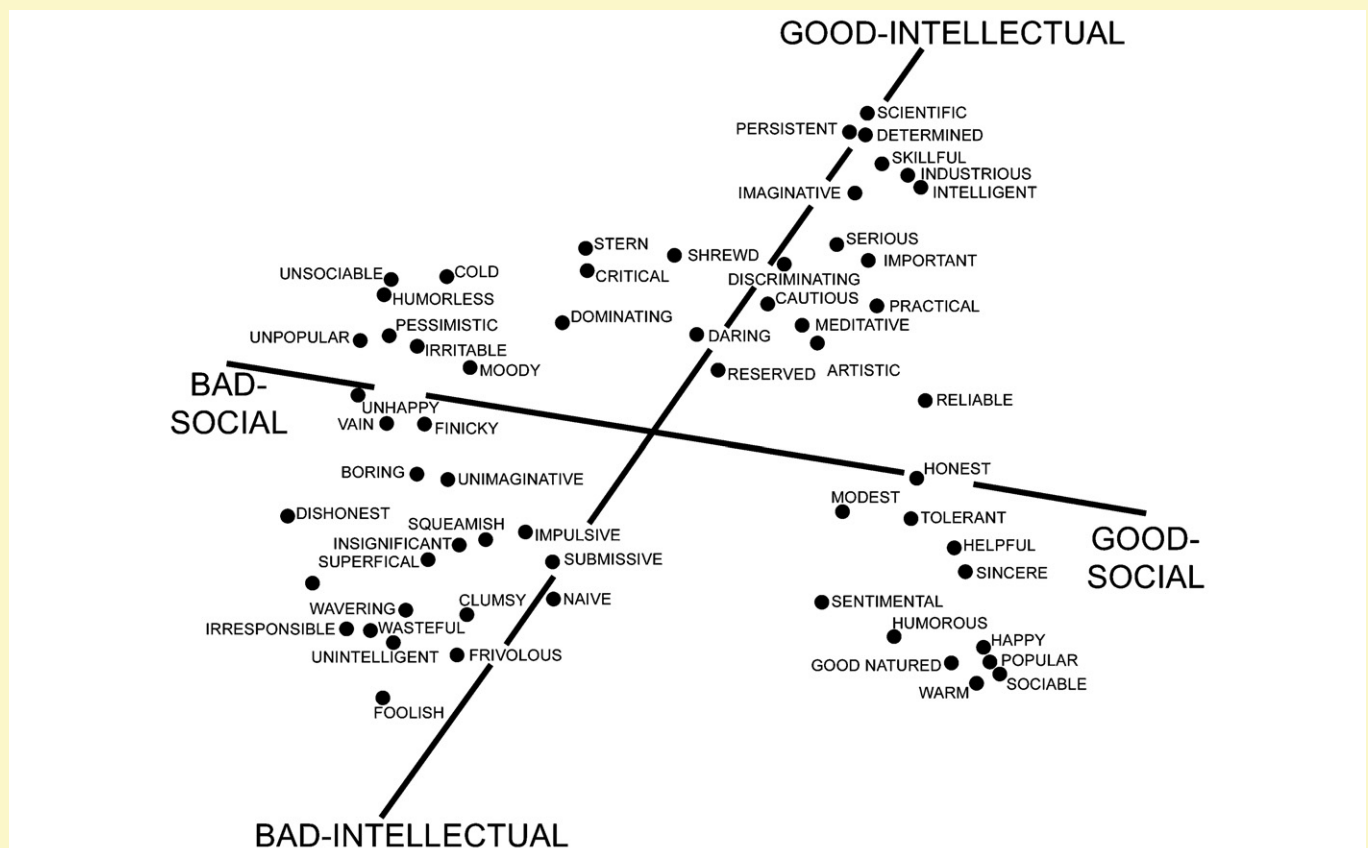

Figure I. Two-dimensional configuration of 60 traits, which shows the best-fitting axes for the properties of social desirability and intellectual desirability. Reproduced, with permission, from Ref. [22].

about the moral–social dimension is more cognitively accessible, more sought by perceivers, more predictive and more heavily weighted in evaluative judgments [5]. The warmth dimension predicts the valence of the interpersonal judgment (i.e. whether the impression is positive or negative), whereas the competence dimension predicts

the extremity of that impression (i.e. *how* positive or *how* negative) [5] (see also Ref. [11]).

Importance of 'other-profitable' traits

Moral–social traits facilitate or hinder other people, whereas competence traits facilitate or hinder mainly the self. The

moral–social, 'other-profitable' traits include kind, honest and aggressive (which is a negative 'other-profitable' trait) because they immediately affect people around the judged person. 'Self-profitable' traits include competence, intelligence and efficiency because they directly and unconditionally affect the possessor's chance of achieving personal goals (e.g. Ref. [9]). In a study that examined 200 trait terms, from a dozen dimensions (including controllability, temporal stability, situational stability and behavioral range), only warmth and competence predicted global evaluations (accounting for 97% of the variance). However, the β-weight for warmth (other-profitable) traits was larger (0.58) than for competence (self-profitable) traits (0.42) [12]. Thus, warmth assessments are primary, at least from the observer's perspective (B. Wojciszke and A.E. Abele, unpublished).

Rapidity of warmth judgments

Cognitively, people are more sensitive to warmth information than to competence information. In lexical decision tasks that control for word length, social perceivers identify warmth-related trait words faster than they identify competence-related trait words [13]. When judging faces after an exposure time of 100 ms, social perceivers judge trustworthiness most reliably, followed by competence [14]. Reliability is calculated by measuring the correlation between time-constrained and time-unconstrained judgments of the same faces. It is striking that people make these judgments in just a fraction of a second, with moral–social judgments occurring first.

Perceivers and situations moderate the primacy of warmth

The priority for detecting warmth over competence, although robust, is stronger for some kinds of perceivers than others. In particular, women, whose traditional gender roles emphasize communal (warmth) over agentic (competence) traits [15], show a stronger priority for detecting warmth [12]. Communal traits traditionally affect women's lives more, whereas competence traits traditionally affect men relatively more [15]. In parallel, collectivist orientations emphasize the social–moral dimension, whereas individualist orientations emphasize the competence dimension [16]. Liking depends on warmth (communion), and respect depends on competence (agency) (A.E. Abele, B. Wojciszke and W. Baryla, unpublished).

Similarly, the relative accessibility of the two dimensions is moderated by the situation. Depending on the primed context, people construe some ambiguous social behaviors in either warmth or competence terms (e.g. tutoring a friend, avoiding a car accident, failing to cheer up a sibling and leaving a meeting). On reading a series of such behaviors, undergraduates interpret them in competence terms if the actions are framed from the actor's (self-related, individualist) perspective and in warm–moral terms if framed from the observer's (other-related, collectivist) perspective [6] (B. Wojciszke and A.E. Abele, unpublished).

Diagnosticity of warmth and competence information: positive and negative

Social perceivers engage a complex calculus regarding relative diagnosticity of the two fundamental dimensions [17–19]. They process positive–negative warmth information and positive–negative competence information asymmetrically, but in opposite ways [17]. Perceivers sensitively heed information that disconfirms, rather than confirms, the other person's warmth [17–20]. This sensitivity reflects concerns about the other person's intentions or motives [10]. To be perceived as warm, a person must adhere to a small range of moral–sociable behavior; a negative deviation eliminates the presumption of morality–warmth and is attributed to the person's (apparently deceptive or mean) disposition. By contrast, a person who is perceived as unfriendly might sometimes behave in moral–sociable ways, but the person will continue to be perceived as unfriendly and untrustworthy; positive deviations are explained by situational demands – even evil people can be nice when it matters to them. In other words, mean and untrustworthy behavior is more diagnostic because it can only be attributed to the other person's disposition, not to social demands. Perceivers interpret warm behavior as controllable, socially cued and, thus, non-diagnostic.

By contrast, perceivers presume that competent behavior is not under immediate personal control. Hence, competence is asymmetrical in a different way from warmth. A person who is perceived as competent might behave competently most of the time, and a few incompetent behaviors do not undermine the perception of general competence (consider the absent-minded professor). However, a person who is perceived as incompetent, and presumably lacks the ability, can never behave competently without challenging the perceived incompetence. Therefore, for competence, positive (compared with negative) behavior is more diagnostic: competence is usually attributed to the other person's abilities, not to social demands.

Sometimes the dimensions combine: competent behavior is particularly diagnostic when the other person is perceived as immoral–unsociable; the competence of an enemy potentially has greater consequences than the competence of a friend [9]. Thus, asymmetries in the processing of positive–negative warmth and competence information can depend on the relative diagnosticity for personality impressions [18–21].

In sum, although both dimensions are fundamental to social perception, warmth judgments seem to be primary, which reflects the importance of assessing other people's intentions before determining their ability to carry out those intentions. This demonstrates a sensitivity to potential threats, which aids survival in all organisms.

Individual versus group perception

Although warmth and competence are separate dimensions [22,23], when people judge individuals, the two dimensions often correlate positively (although modestly) in the well-known halo effect [22,24]: people expect isolated individuals to be evaluatively consistent [25]. However, when people judge social groups, warmth and competence often correlate negatively: many groups are judged as high on one dimension and low on the other, which has important implications for affective and behavioral reactions [26–28].

People ask the same warmth and competence questions of societal ingroups and outgroups as they do of individuals,

which creates predictable stereotypes, emotional prejudices and discriminatory tendencies. (By convention, social psychologists refer to a perceiver's own group as the ingroup and all others as outgroups [29].) The types of bias against outgroups differ depending on the group and its perceived relationship to other groups in society.

Stereotype content model

The two-dimensional warmth-by-competence space depicts one societal ingroup and three kinds of outgroups that are recognizable in all the countries that have been studied (see below). From the societal perspective, certain groups are prototypes or, in sociological terms, reference groups. For example, in the USA, at the present time, middle-class people, Christian people, heterosexual people and US citizens all are societal ingroups. People rate these groups as high on both warmth and competence, and they express pride and admiration for them [28,30,31] (Figure 1).

Prejudice is not simply antipathy

Lay people and psychologists have long viewed outgroup prejudice as antipathy [32], whereby societal outgroups are stereotypically neither warm nor competent, but hostile, untrustworthy, stupid and unmotivated. In the USA, these groups are reported to include poor white people, poor black people, welfare recipients, homeless people, drug addicts and undocumented migrants [28,30,31,33]. These groups reportedly elicit contempt and disgust more than all other groups. On viewing photographs of apparently homeless or addicted individuals, perceivers show neural activation in the insula, which is consistent with disgust. Furthermore, areas that are normally activated on viewing or thinking about other people (e.g. the medial prefrontal cortex) show significantly less activation to these outgroups, as if people perceive them as less than human [34].

Ambivalent prejudices

Although some outgroups are perceived negatively on both warmth and competence, others are perceived ambivalently (high on one dimension and low on the other). Most societal outgroups fall into these previously ignored combinations [30,31,35]. US data show that people who are older, physically disabled or mentally disabled are viewed as warm but incompetent. These groups elicit pity and sympathy [28,30,31,36], which are inherently ambivalent emotions that communicate subordinate status but paternalistic positivity [37].

Other groups are viewed as competent but cold (and untrustworthy). In the USA, these currently include rich people, Asian people, Jewish people, female professionals and minority professionals [28,30,31]. These groups elicit envy and jealousy more than other groups. Such resentful emotions are inherently ambivalent because they suggest that the outgroup possesses prized abilities but that their intentions are suspect.

The US evidence for these four combinations of warmth and competence includes 'convenience' samples of undergraduates, their parents and retirement communities [30], and also a representative sample survey of US adults [31]. The four types of outgroups also seem to fit studies of ethnic stereotypes that have persisted since the 1930s (L.M.
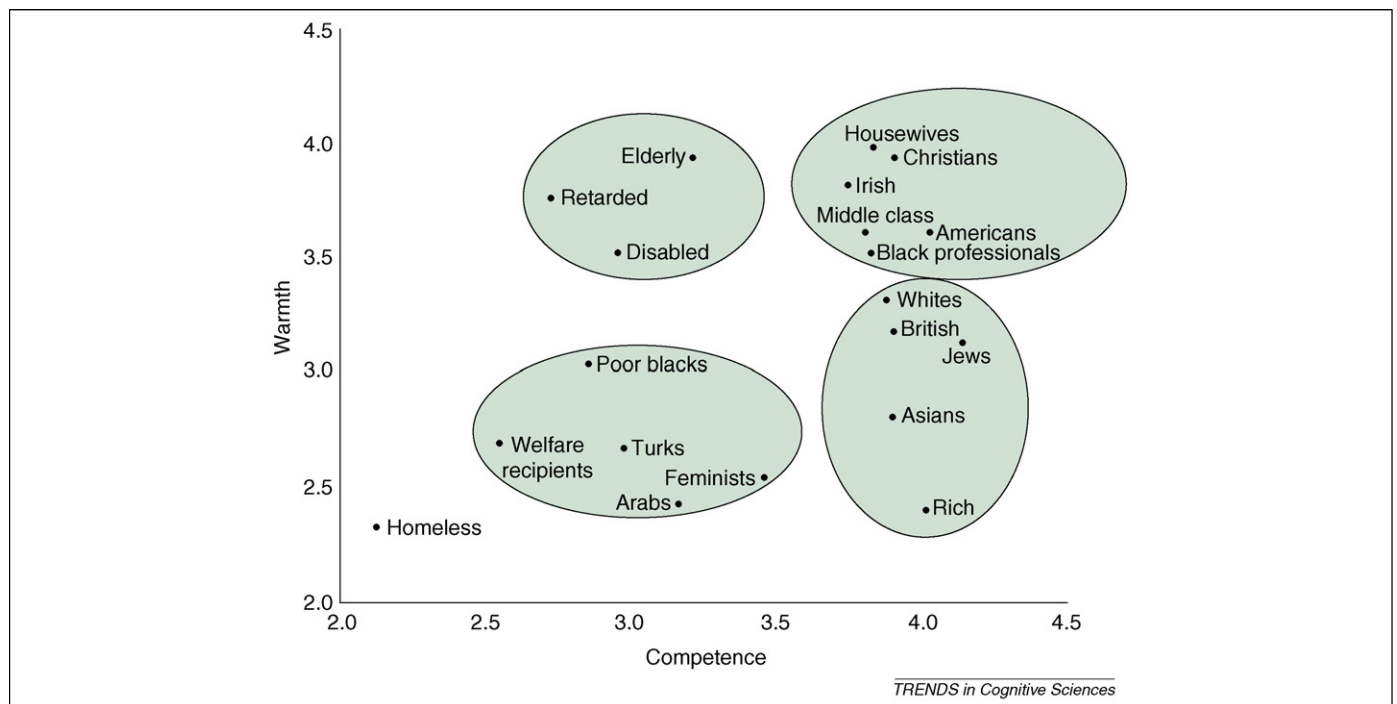

Figure 1. Scatter plot and cluster analysis of competence and warmth ratings for 20 groups. Averaging across US respondents, each group receives warmth (warm, friendly) and competence (competent, capable) scores, which are submitted to cluster analyses to determine number and membership of clusters. Groups near the center of their cluster replicate cluster membership most reliably across studies. Ratings on other variables (emotions, behaviors) cross-validate the cluster solutions. Different group names were used in different studies. Usually, an initial sample of respondents generated group names that were later rated by a second set of respondents on warmth and competence. The 20 names shown here were selected from prior sets and for various theoretical reasons. Warmth and competence were rated on five-point scales. Related to data from Refs [30,31,33,38,39,43]. Reproduced, with permission, from Ref. [31].

Leslie, V.S. Constantine and S.T. Fiske, unpublished) and they fit every society that has been studied so far: 19 nations on 4 continents [35,36,38,39] (I. Anselin and S.T. Fiske, unpublished; A.J.C. Cuddy and S.T. Fiske, unpublished). (To our knowledge, nations studied so far are, in North America, the USA, Costa Rica, Dominican Republic and Mexico; in Europe, Belgium, Bulgaria, France, Germany, Italy, the Netherlands, Norway, Portugal, Spain and the UK; in the Middle East, Israel; in Africa, South Africa; and in Asia, Hong Kong, Japan and South Korea.) In every society studied, poor people are perceived as neither nice nor smart, rich people are perceived as smart but not nice and older people are perceived as nice but not smart. Other societal groups that are local to each culture fit these three classifications. (The one exception is that in Asian cultures, in keeping with modesty norms, people rate societal ingroups neutrally on competence and warmth; however, the other three combinations are fully represented [38]. This demonstrates that outgroup prejudice does not require overt ingroup admiration.)

The warmth-by-competence space also fits in-depth US perceptions of specific US societal subgroups, such as subtypes of older people [40,41], Asian and Asian–American people [42], subgroups of immigrants [33], subtypes of gay men [43], subgroups of women [39,44], people who have distinct mental illnesses (A.M. Russell, S.T. Fiske, G. Moore and D. Thompson, unpublished), European nationalities [38,45–47], enemy outgroups [48], socioeconomic groups [49–51] and speakers of nonstandard dialects [52].

Behavioral consequences
Distinct types of discrimination result from each warmth-by-competence combination, which is captured by the behaviors from intergroup affect and stereotypes (BIAS) map [35] (Figure 2). Being primary, the warmth dimension predicts active behaviors: active facilitation (helping) versus active harming (attacking). Being secondary, the
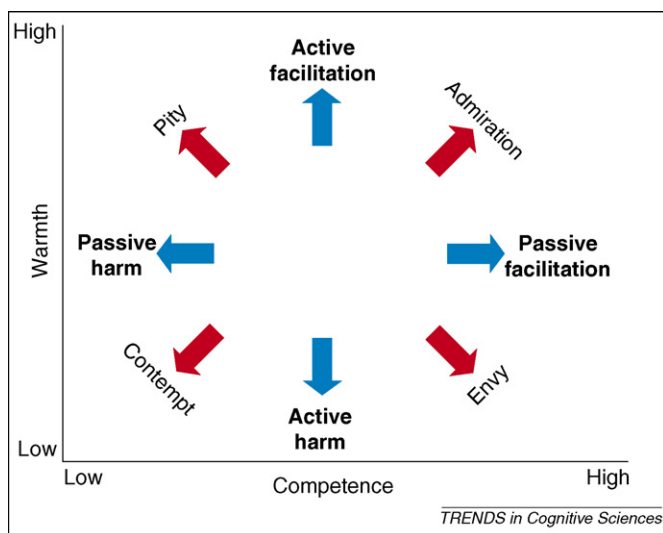

Figure 2. Schematic representation of behaviors from intergroup affect and stereotypes (BIAS) map. Competence and warmth stereotypes are represented along the *x* and *y* axes. Emotions are represented by red arrows on diagonal axes. Thus, groups in different quadrants are rated as receiving one predicted emotional prejudice and two predicted behaviors. Behavioral tendencies are represented by blue arrows on horizontal and vertical axes. Reproduced, with permission, from Ref. [31].

competence dimension predicts passive behaviors: passive facilitation (association) and passive harm (neglect).

The intersections of the two dimensions create unique behavioral profiles that are directed towards each type of outgroup. In the two most straightforward cases, societal ingroups elicit both active and passive facilitation (helping and associating) and the low–low outgroups (e.g. homeless people) receive both kinds of harm (active attacks and passive neglect) [31]. News reports confirm this potentially fatal kind of discrimination.

The mixed combinations are more volatile: pitied groups (e.g. older and disabled people) elicit active helping and passive neglect; for example, institutionalizing older or disabled people actively aids them but socially isolates them. By contrast, envied groups elicit passive association and active harm [31]; for example, neighbors might shop at the stores of entrepreneurial outsiders but, under societal breakdown, they might attack and loot these same shops. Jews during the Holocaust, Koreans in the Los Angeles riots and Chinese in the Indonesian riots all exemplify this unfortunate profile.

What reliably predicts these discriminatory behaviors? In path analyses of representative data from the USA, competence and warmth stereotypes combine to predict emotions, which directly predict behaviors [31]. The proximal cause of these social behaviors is affect, a finding that is reflected in meta-analyses of emotional prejudices and cognitive stereotypes as predictors of discrimination [53–55]. Stereotypes can legitimize antipathy towards outgroups [49,50,56,57]. However, the social structure creates these relationships of antipathy and stereotyping, as we show next.

Antecedents of stereotypes, emotions and behaviors
Groups often compete with each other or at least do not facilitate each other's goals. Definitions of what constitutes a group often include shared goals, which presumably differ from the goals of other groups. When perceivers view the goals of an outgroup as differing from or conflicting with goals of the ingroup, they ascribe negative traits and experience negative emotions towards the outgroup [56]. Thus, when a group explicitly competes with the ingroup or exploits the ingroup, its intent is seen as unfriendly and untrustworthy (i.e. not warm). By contrast, when a group cooperates with or does not hinder the ingroup, then their intent is seen as friendly and trustworthy (i.e. warm). This can be viewed as perceived threat, over competition for resources.

As this theory predicts, the perceived warmth and interdependence (cooperation–competition) of groups are negatively correlated (on average, −0.52 across groups and −0.27 across individuals) across US, Western European and Asian samples [30,31,38]. The items that measure competition include power and resource tradeoffs (if one group gains power, then other groups lose power; resources that go to one group take resources away from the rest of society).

The other dimension, competence, results from judged status. To the extent that people justify hierarchical systems [58] or believe in a just world [57], they believe that groups get what they deserve. People assume that

high- versus low-status groups merit their positions because they are, respectively, more versus less competent.

Of the 19 nations we have studied, the status–competence correlations average 0.94 across groups and 0.77 across individuals [30,31,38], which suggests that these constructs are, effectively, identical. Yet the status measure includes prestigious jobs (which potentially could result from advantageous birth, connections or nepotism) and economic success (which potentially could result from luck or inheritance); the status measure is demographic, whereas the competence measure comprises traits. However, instead of resentment towards the privileged and sympathy for the underdog, on average, people endorse the apparent meritocracy and infer that (for groups) high status invariably reflects competence. However, people vary ideologically; people who endorse group hierarchies or who believe in a just world show higher status–competence correlations for perceptions of generic individuals [59].

Evidence for these social structural predictors (status and interdependence) is not only correlational, but also causal, based on experimental investigations of intergroup perception. When US citizens rate hypothetical groups that vary in ascribed status (P. Caprariello, A.J.C. Cuddy and S.T. Fiske, unpublished) or guess about unseen people living in expensive versus inexpensive houses [59], they infer the competence of the groups and individuals that are involved. Inter-nation perceptions show similar findings [48,60].

Returning to individual-person perception, new findings suggest interpersonal parallels to these intergroup predictors. Individuals who are arbitrarily placed in competition or cooperation respectively dislike or like each other; likewise, random assignment to status determines respect or disrespect (A.M. Russell and S.T. Fiske, unpublished). Like groups, individuals differentiate upward from downward status and contrast competition with assimilation [61].

Summary

Warmth and competence are reliably universal dimensions of social judgment across stimuli, cultures and time. The consistency with which these dimensions appear might reflect the answers to two basic survival questions: first, and crucially, does the other person or group intend to harm or help me (or us)? Secondarily, does the other have the ability to enact those intentions? If these dimensions do reflect survival value, warmth and competence are not merely psychometric curiosities but enduring, fundamental and (arguably) evolved aspects of social perception. Furthermore, how individuals and groups are perceived on these dimensions results from structural relationships. Interdependence predicts perceived warmth, and status predicts perceived competence. Particular combinations of these perceived dimensions have distinct emotional and behavioral consequences. This is a particularly pertinent issue in terms of group-based prejudices. Typically, group stereotypes appear high on one dimension and low on the other; the ensuing ambivalent affect and volatile behavior potentially endanger constructive intergroup relationships.

References

1 Abelson, R.P. *et al.* (1982) Affective and semantic components in political person perception. *J. Pers. Soc. Psychol.* 42, 619–630
2 Kinder, D.R. *et al.* (1980) Presidential prototypes. *Polit. Behav.* 2, 315–338
3 Wojciszke, B. and Klusek, B. (1996) Moral and competence-related traits in political perception. *Pol. Psychol. Bull.* 27, 319–324
4 Chemers, M.M. (1997) *An Integrative Theory of Leadership,* Lawrence Erlbaum Associates
5 Wojciszke, B. *et al.* (1998) On the dominance of moral categories in impression formation. *Pers. Soc. Psychol. Bull.* 24, 1245–1257
6 Wojciszke, B. (1994) Multiple meanings of behavior: construing actions in terms of competence or morality. *J. Pers. Soc. Psychol.* 67, 222–232
7 Wojciszke, B. (2005) Affective concomitants of information on morality and competence. *Eur. Psychol.* 10, 60–70
8 Cacioppo, J.T. *et al.* (1997) Beyond bipolar conceptualizations and measures: the case of attitudes and evaluative space. *Pers. Soc. Psychol. Rev.* 1, 3–25
9 Peeters, G. (2001) From good and bad to can and must: subjective necessity of acts associated with positively and negatively valued stimuli. *Eur. J. Soc. Psychol.* 31, 125–136
10 Reeder, G.D. *et al.* (2002) Inferences about the morality of an aggressor: the role of perceived motive. *J. Pers. Soc. Psychol.* 83, 789–803
11 Wojciszke, B. *et al.* (1993) Effects of information content and evaluative extremity on positivity and negativity biases. *J. Pers. Soc. Psychol.* 64, 327–336
12 Wojciszke, B. *et al.* (1998) Moral competence-related traits: how do they differ? *Pol. Psychol. Bull.* 29, 283–294
13 Ybarra, O. *et al.* (2001) Young and old adults' concerns about morality and competence. *Motiv. Emot.* 25, 85–100
14 Willis, J. and Todorov, A. (2006) First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17, 592–598
15 Abele, A.E. (2003) The dynamics of masculine–agentic and feminine–communal traits: findings from a prospective study. *J. Pers. Soc. Psychol.* 85, 768–776
16 Wojciszke, B. (1997) Parallels between competence- versus morality-related traits and individualistic versus collectivistic values. *Eur. J. Soc. Psychol.* 27, 245–256
17 Singh, R. and Teoh, J.B.P. (2000) Impression formation from intellectual and social traits: evidence for behavioural adaptation and cognitive processing. *Br. J. Soc. Psychol.* 39, 537–554
18 Ybarra, O. and Stephan, W.G. (1999) Attributional orientations and the prediction of behavior: the attribution–prediction bias. *J. Pers. Soc. Psychol.* 76, 718–727
19 Skowronski, J.J. and Carlston, D.E. (1987) Social judgment and social memory: the role of cue diagnosticity in negativity, positivity, and extremity biases. *J. Pers. Soc. Psychol.* 52, 689–699
20 Tausch, N. *et al.* The confirmability and disconfirmability of trait concepts revisited: does content matter? *J. Pers. Soc. Psychol.* (in press)
21 Fiske, S.T. (1980) Attention and weight in person perception: the impact of negative and extreme behavior. *J. Pers. Soc. Psychol.* 38, 889–906
22 Rosenberg, S. *et al.* (1968) A multidimensional approach to the structure of personality impressions. *J. Pers. Soc. Psychol.* 9, 283–294
23 Zanna, M. and Hamilton, D. (1972) Attribute dimensions and patterns of trait inferences. *Psychon. Sci.* 27, 353–354
24 Judd, C.M. *et al.* (2005) Fundamental dimensions of social judgment: understanding the relations between judgments of competence and warmth. *J. Pers. Soc. Psychol.* 89, 899–913
25 Fiske, S.T. and Taylor, S.E. *Social Cognition* (3rd edn), McGraw-Hill (in press)
26 Fiske, S.T. (1998) Stereotyping, prejudice, and discrimination. In *The Handbook of Social Psychology* (4th edn) (Gilbert, D.T. *et al.*, eds), pp. 357–411, McGraw-Hill
27 Yzerbyt, V. *et al.* (2005) Not competent but warm…really? Compensatory stereotypes in the French-speaking world. *Group Process. Intergroup Relat.* 8, 291–308
28 Fiske, S.T. *et al.* (1999) (Dis)respecting versus (dis)liking: status and interdependence predict ambivalent stereotypes of competence and warmth. *J. Soc. Issues* 55, 473–491

29 Brewer, M.B. and Brown, R.J. (1998) Intergroup relations. In *The Handbook of Social Psychology* (4th edn) (Gilbert, D.T. *et al.*, eds), pp. 554–594, McGraw–Hill

30 Fiske, S.T. *et al.* (2002) A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902

31 Cuddy, A.J.C. *et al.* The BIAS map: behaviors from intergroup affect and stereotypes. *J. Pers. Soc. Psychol.* (in press)

32 Allport, G.W., ed. (1954) *The Nature of Prejudice*, Addison-Wesley

33 Lee, T.L. and Fiske, S.T. (2006) Not an outgroup, but not yet an ingroup: immigrants in the stereotype content model. *Int. J. Intercult. Relat.* 30, 751–768

34 Harris, L.T. and Fiske, S.T. (2006) Dehumanizing the lowest of the low: neuro-imaging responses to extreme outgroups. *Psychol. Sci.* 17, 847–853

35 Cuddy, A.J.C. *et al.* Competence and warmth as universal trait dimensions of interpersonal and intergroup perception: the stereotype content model and the BIAS map. In *Advances in Experimental Social Psychology* (Zanna, M.P., ed.), Academic Press (in press)

36 Fiske, S.T. and Cuddy, A.J.C. (2006) Stereotype content and relative group status across cultures. In *Social Comparison Processes and Levels of Analysis: Understanding Culture, Intergroup Relations and Cognition* (Guimond, S., ed.), pp. 249–263, Cambridge University Press

37 Glick, P. and Fiske, S.T. (2001) Ambivalent sexism. In *Advances in Experimental Social Psychology* (Vol. 33) (Zanna, M.P., ed.), pp. 115–188, Academic Press

38 Cuddy, A.J.C. *et al.* Is the stereotype content model culture-bound? A cross-cultural comparison reveals systematic similarities and differences. *Br. J. Soc. Psychol.* (in press)

39 Eckes, T. (2002) Paternalistic and envious gender stereotypes: testing predictions from the stereotype content model. *Sex Roles* 47, 99–114

40 Cuddy, A.J.C. and Fiske, S.T. (2002) Doddering, but dear: process, content, and function in stereotyping of older persons. In *Ageism* (Nelson, T.D., ed.), pp. 3–26, MIT Press

41 Cuddy, A.J.C. *et al.* (2005) This old stereotype: the pervasiveness and persistence of the elderly stereotype. *J. Soc. Issues* 61, 265–283

42 Lin, M.H. *et al.* (2005) Stereotype content model explains prejudice for an envied outgroup: scale of anti-Asian American stereotypes. *Pers. Soc. Psychol. Bull.* 31, 34–47

43 Clausell, E. and Fiske, S.T. (2005) When do the parts add up to the whole? Ambivalent stereotype content for gay male subgroups. *Soc. Cogn.* 23, 157–176

44 Cuddy, A.J.C. *et al.* (2004) When professionals become mothers, warmth doesn't cut the ice. *J. Soc. Issues* 60, 701–718

45 Linssen, H. and Hagendoorn, L. (1994) Social and geographical factors in the explanation of the content of European nationality stereotypes. *Br. J. Soc. Psychol.* 33, 165–182

46 Phalet, K. and Poppe, E. (1997) Competence and morality dimensions in national and ethnic stereotypes: a study in six eastern-European countries. *Eur. J. Soc. Psychol.* 27, 703–723

47 Petkova, K. and Todorov, V. (2002) Bulgarian national stereotypes and national identity. *Sociol. Probl.* 34, 115–128

48 Alexander, M.G. *et al.* (1999) Images and affect: a functional analysis of out-group stereotypes. *J. Pers. Soc. Psychol.* 77, 78–93

49 Kay, A.C. and Jost, J.T. (2003) Complementary justice: effects of 'poor but happy' and 'poor but honest' stereotype exemplars on system justification and implicit activation of the justice motive. *J. Pers. Soc. Psychol.* 85, 823–837

50 Kay, A.C. *et al.* (2004) Victim derogation and victim enhancement as alternate routes to system justification. *Psychol. Sci.* 16, 240–246

51 Johannesen-Schmidt, M. and Eagly, A. (2002) Diminishing returns: the effects of income on the content of stereotypes of wage earners. *Pers. Soc. Psychol. Bull.* 28, 1538–1545

52 Ruscher, J.B. (2001) *The Social Psychology of Prejudiced Communication,* Guilford Press

53 Dovidio, J.F. *et al.* (1996) Stereotyping, prejudice, and discrimination: another look. In *Stereotypes and Stereotyping* (Macrae, C.N. *et al.*, eds), pp. 276–319, Guilford Press

54 Esses, V.M. and Dovidio, J.F. (2002) The role of emotions in determining willingness to engage in intergroup contact. *Pers. Soc. Psychol. Bull.* 28, 1202–1214

55 Talaska, C.A. *et al.* Legitimating racial discrimination: a meta-analysis of the racial attitude–behavior literature shows that emotions, not beliefs, best predict discrimination. *Soc. Justice Res.* (in press)

56 Fiske, S.T. and Ruscher, J.B. (1993) Negative interdependence and prejudice: whence the affect? In *Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception* (Mackie, D.M. and Hamilton, D.L., eds), pp. 239–268, Academic Press

57 Lerner, M.J. and Miller, D.T. (1978) Just world research and the attribution process: looking back and ahead. *Psychol. Bull.* 85, 1030–1051

58 Jost, J.T. and Banaji, M.R. (1994) The role of stereotyping in system–justification and the production of false consciousness. *Br. J. Soc. Psychol.* 33, 1–27

59 Oldmeadow, J. and Fiske, S.T. Ideology moderates status = competence stereotypes: roles for belief in a just world and social dominance orientation. *Eur. J. Soc. Psychol.* (in press)

60 Poppe, E. (2001) Effects of changes in GNP and perceived group characteristics on national and ethnic stereotypes in central and eastern Europe. *J. Appl. Soc. Psychol.* 31, 1689–1708

61 Smith, R.H. (2000) Assimilative and contrastive emotional reactions to upward and downward social comparisons. In *Handbook of Social Comparison: Theory and Research* (Suls, J. and Wheeler, L., eds), pp. 173–200, Kluwer Academic/Plenum

62 Asch, S.E. (1946) Forming impressions of personality. *J. Abnorm. Soc. Psychol.* 42, 258–290

63 Bales, R.F. (1950) A set of categories for the analysis of small group interaction. *Am. Sociol. Rev.* 15, 257–263

64 Bales, R.F., ed. (1999) *Social Interaction Systems: Theory and Measurement*, Transaction Press

Home   Read   Sign in

Search in book …

PRINCIPLES OF SOCIAL PSYCHOLOGY – 1ST INTERNATIONAL EDITION

CONTENTS

5. Perceiving Others

Initial Impression Formation

Learning Objectives

1. Describe how people use behaviors and traits to form initial perceptions of others.

2. Explore research about forming impressions from thin slices of information.

3. Summarize the role of nonverbal behaviors in person perception.

4. Review research about detecting deception.

People are often very skilled at **person perception**—*the process of learning about other people*—and our brains are designed to help us judge others efficiently (Haselton & Funder, 2006; Macrae & Quadflieg, 2010). Infants
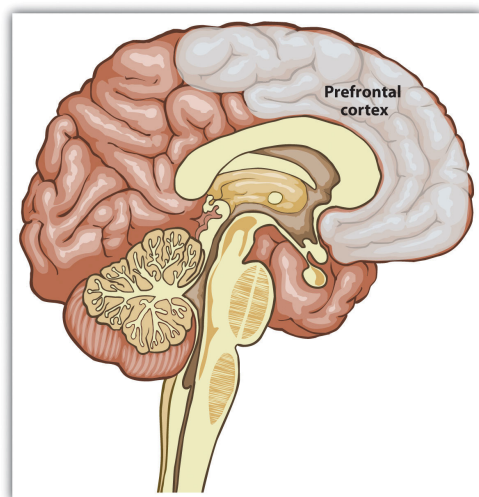
remember a potentially unlimited number of people as we navigate our social environments [Increase Font Size] Gobbini, 2000), and we form impressions of those others quickly and without much effort (Carlston & Skowronski, 2005; Fletcher-Watson, Findlay, Leekam, & Benson, 2008). Furthermore, our first impressions are, at least in some cases, remarkably accurate (Ambady, Bernieri, & Richeson, 2000).

Recent research is beginning to uncover the areas in our brain where person perception occurs. In one relevant study, Mason and Macrae (2004) used functional magnetic resonance imaging (fMRI) scans to test whether people stored information about other people in a different location in the brain than where they stored information about animals, and they found that this was the case. Specific areas of the prefrontal cortex were found to be more active when people made judgments about people rather than dogs (Figure 5.2).

Figure 5.2 Recent advances in neuroimaging techniques have provided information about the brain structures that are involved in person perception. The prefrontal cortex shows strong activation when we are thinking about another person. Data are from Mason, Banfield, and Macrae (2004). Mason, M. F., & Macrae, C. N. (2004). Categorizing and individuating others: The neural substrates of person perception. Journal of Cognitive Neuroscience, 16(10), 1785–1795. doi: 10.1162/0898929042947801

Learning about people is a lot like learning about any other object in our environment, with one major exception. With an object, there is no interaction: we learn about the characteristics of a car or a cell phone, for example, without any concern that the car or the phone is learning about us. It is a one-way process. With people, in contrast, there is a two-way social process: just as we are learning about another person, that person is learning about us, or potentially attempting to keep us from accurately perceiving him or her. For instance, research has found that when other people are looking directly at us, we process their features more fully and faster, and we remember them better than when the same people are not looking at us (Hood & Macrae, 2007).

In the social dynamic with others, then, we have two goals: first, we need to learn about them, and second, we want them to learn about us (and, we hope, like and respect us). Our focus here is on the former process—how we make sense of other people. But remember that just as you are judging them, they are judging you.

We have seen in the chapter, "The Self", that when people are asked to describe themselves, they generally do so

ers. In this section, we will review how we initially use the physical features and social categ **Increase Font Size** others (e.g., male or female, race, and ethnicity) to form judgments and then will focus on the role of personality traits in person perception.

Figure 5.3 One of the important tasks of everyday life is to form judgments about other people. Source: Terrorist Disguised as a Woman by Israel Defense Forces (http://commons.wikimedia.org/wiki/File:Flickr_-_Israel_Defense_Forces_-_Terrorist_Disguised_as_a_Woman.jpg) used under CC BY SA 3.0 license (http://creativecommons.org/licenses/by-sa/3.0/deed.en); You are not listening! by Jesslee Cuizon (http://commons.wikimedia.org/wiki/File:You're_not_listening!.jpg) used under CC BY 2.0 (http://creativecommons.org/licenses/by/2.0/deed.en); Family Party by Fairfax County (https://www.flickr.com/photos/fairfaxcounty/8617461034) used under CC BY NC ND 2.0 license (https://creativecommons.org/licenses/by-nc-nd/2.0/); North Charleston Police Officers by North Charleston (https://www.flickr.com/photos/northcharleston/8960603856) used under CC BY SA 2.0 license (https://creativecommons.org/licenses/by-sa/2.0/).

Research Focus

Forming Impressions from Thin Slices

Although it might seem surprising, social psychological research has demonstrated that at least in some limited situations, people can draw remarkably accurate conclusions about others on the basis of very little data, and that they can do this very quickly (Rule & Ambady, 2010; Rule, Ambady, Adams, & Macrae, 2008; Rule, Ambady, & Hallett, 2009).

Ambady and Rosenthal (1993) made videotapes of six female and seven male graduate students while they were teaching an undergraduate course. The courses covered diverse areas of the college curriculum, including humanities, social sciences, and natural sciences. For each instructor, three 10-second video clips were taken—10 seconds from the first 10 minutes of the class, 10 seconds from the middle of the class, and 10 seconds from the last 10 minutes of the class.

rating. Ambady and her colleagues then compared the ratings of the instructors made bIncrease Font Size
pants who had seen the instructors for only 30 seconds with the ratings of the same instructors that had
been made by actual students who had spent a whole semester with the instructors and who had rated
them at the end of the semester on dimensions such as "the quality of the course section" and "the sec-
tion leader's performance." The researchers used the Pearson correlation coefficient to make the com-
parison (remember that correlations nearer +1.0 or −1.0 are stronger). As you can see in the following
table, the ratings of the participants and the ratings of the students were highly positively correlated.

Table 5.1 Forming Accurate Impressions in Only 30 Seconds

Correlations of Molar Nonverbal Behaviors with College Teacher Effectiveness Ratings (Student Ratings)	
Variable	*r*
Accepting	.50
Active	.77**
Attentive	.48
Competent	.56*
Confident	.82***
Dominant	.79**
Empathic	.45
Enthusiastic	.76**
Honest	.32
Likable	.73**
(Not) Anxious	.26
Optimistic	.84***
Professional	.53
Supportive	.55*
Warm	.67*
Global Variable	.76**
***p*<.05. ***p*<.01. ****p*<.001. Data are from Ambady and Rosenthal (1993). Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64*(3), 431–441.**	

If the finding that we can make accurate judgments about other people in only 30 seconds surprises
you, then perhaps you will be even more surprised to learn that we do not even need that much time.
Willis and Todorov (2006) found that even a tenth of a second was enough to make judgments that cor-
related highly with the same judgments made by other people who were given several minutes to make
the judgments. Other research has found that we can make accurate judgments in seconds or even mil-
liseconds about, for instance, the personalities of salespersons (Ambady, Krabbenhoft, & Hogan,
2006) and even whether or not a person is prejudiced (Richeson & Shelton, 2005).

had run against each other in previous elections for the U.S. Senate and House of Repr Increase Font Size ticipants saw only the faces of the candidates, and they saw them in some cases for only one second. Their task was to judge which person in of each pair was the most competent. Todorov and colleagues (2005) found that these judgments predicted the actual result of the election; in fact, 68% of the time the person judged to have the most competent face won.

Rule and Ambady (2010) showed that perceivers were also able to accurately distinguish whether people were Democrats or Republicans based only on photos of their faces. Republicans were perceived as more powerful than Democrats, and Democrats were perceived as warmer than Republicans. Further, Rule, Ambady, Adams, and Macrae (2008) found that people could accurately determine the sexual orientation of faces presented in photos (gay or straight) based on their judgments of what they thought "most people" would say. These findings have since been replicated across different cultures varying in their average acceptance of homosexuality (Rule, Ishii, Ambady, Rosen, & Hallett, 2011).

Taken together, these data confirm that we can form a wide variety of initial impressions of others quickly and, at least in some cases, quite accurately. Of course, in these situations, the people who were being observed were not trying to hide their personalities from the observers. As we saw in Chapter 3, people often use strategic self-presentation quite skillfully, which further complicates the person perception process.

Nonverbal Behavior

One way that the participants in the studies described above may have been able to form such accurate impressions of instructors on the basis of such little information was by viewing their nonverbal behavior. **Nonverbal behavior** is *any type of communication that does not involve speaking, including facial expressions, body language, touching, voice patterns, and interpersonal distance.* Nonverbal behaviors are used to reinforce spoken words (Hostetter, 2011) but also include such things as interpersonal distance (how far away from you the other person stands), tone of voice, eye gaze, and hand gestures and body positions (DePaulo et al., 2003).

The ability to decode nonverbal behavior is learned early, even before the development of language (Walker-Andrews, 2008). We tend to like people who have a pleasant tone of voice and open posture, who stand an appropriate distance away from us, and who look at and touch us for the "right" amount of time—not too much or too little. And, of course, behavior matters; people who walk faster are perceived as happier and more powerful than those who walk more slowly (Montepare & Zebrowitz-McArthur, 1988). (For more insight into the relationship between nonverbal communication and success, see social psychologist Amy Cuddy's fascinating TED Talk here: http://www.ted.com/talks/amy_cuddy_your_body_language_shapes_who_you_are).

recognize the behavior of others from these minimal displays (Clarke, Bradshaw, Field, Ham[**Increase Font Size**]
Johnson, Gill, Reichman, & Tassinary, 2007; Heberlein, Adolphs, Tranel, & Damasio, 2004; Figure 5.4 "Point-Light Displays"). People can also determine personality by tone of voice provided by degraded and incomprehensible speech (Ambady, Krabbenhoft, & Hogan, 2006).

Figure 5.4 Point-Light Displays

People can accurately detect behaviors, emotions, and traits from point-light displays. You might want to try your skills here: http://astro.temple.edu/~tshipley/mocap/dotMovie.html.

Although they may be pretty good at it in some cases, people are often not aware of their ability to make accurate judgments. Rule, Ambady, Adams, and Macrae (2008) found that even though the participants in their research were quite accurate in their perceptions, they could not articulate how they made their judgments. They claimed that they were "just guessing" and could hardly believe that they were getting the judgments right. These results suggest that they were made without any conscious awareness on the part of the judgers. Furthermore, the participants' judgments of their own accuracy were not generally correlated with their actual accurate judgments.

Figure 5.5 Nonverbal behaviors are an important form of communication—and they are particularly important in expressing our liking of, and caring for, others. Source: Touch by eternal sunshine (https://www.flickr.com/photos/yugandhar/997464862/) used under CC BY NC SA 2.0 license (https://creativecommons.org/licenses/by-nc-sa/2.0/); 02710009 by IAEA (https://www.flickr.com/photos/iaea_imagebank/8388691703/) Imagebank used under CC BY SA 2.0 (https://creativecommons.org/licenses/by-sa/2.0/); Mother and Daughter by imagebang

The particular nonverbal behaviors that we use, as well as their meanings, are determined by social norms, and these norms may vary across cultures. For example, people who live in warm climates nearer the equator use more nonverbal communication (e.g., talking with their hands or showing strong facial expressions) and are more likely to touch each other during conversations than people who live in colder climates nearer Earth's poles (Manstead, 1991; Pennebaker, Rime, & Blankenship, 1996). And the appropriate amount of personal space to keep between ourselves and others also varies across cultures. In some cultures—for instance, those of South American countries—it is appropriate to stand very close to another person while talking to him or her; in other cultures—for example, in the United States and Western Europe—more interpersonal space is the norm (Knapp & Hall, 2006). The appropriate amount of eye contact with others is also determined by culture. In Latin America, it is appropriate to lock eyes with another person, whereas in Japan, people more often try to avoid eye contact.

Although nonverbal behaviors can be informative during the initial stages of person perception, they are limited in what they can convey. In general, they communicate our own status or dominance (self-concern) as well as our interest in or liking of another (other-concern). If we notice that someone is smiling and making eye contact with us while leaning toward us in conversation, we can be pretty sure that he or she likes us. On the other hand, if someone frowns at us, touches us inappropriately, or moves away when we get close, we may naturally conclude that he or she does not like us.

We may also use nonverbal behaviors to try out new situations: If we move a little closer and look at someone a bit longer, we communicate our interest. If these responses are reciprocated by the other person, that can indicate that he or she likes us, and we can move on to share other types of information. If the initial nonverbal behaviors are not reciprocated, then we may conclude that the relationship may not work out and we can withdraw before we go "too far."

Nonverbal behavior provides different information than verbal behavior because people frequently say one thing and do another. Perhaps you remember being really angry at someone but not wanting to let on that you were mad, so you tried to hide your emotions by not saying anything. But perhaps your nonverbal behavior eventually gave you away to the other person: although you were trying as hard as you could not to, you just looked angry. We frequently rely more on nonverbal than on verbal behavior when messages are contradictory. One reason for this is that we know that it is relatively easy to monitor our verbal behavior but harder to control the nonverbal. However, we expect that people who need to deceive others—for instance, good poker players—are able to monitor their nonverbal behavior better than most people, making it difficult to get a good read on them.

Because we use nonverbal behaviors so frequently in our social interactions, we are fluent readers of them. We also realize that we can better communicate with others when we use them. Indeed, it is difficult to communicate accurately when we cannot express ourselves nonverbally (Krauss, Chen, & Chawla, 1996). You probably have noticed this yourself. If you e-mail or text a message to your friend, for instance, you need to be careful about using sarcasm because he or she might misinterpret your meaning. Because nonverbal information is so important, we quickly learned to incorporate it, in the form of emoticons, in our text messages (Figure 5.6).

Figure 5.6 Emoticons are a type of nonverbal behavior for electronic messages. Source: Emoticons by Gustavo26776 (http://wikimediafoundation.org/wiki/File:Emoticons.gif) used under CC BY SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0/deed.en)

Detecting Danger by Focusing on Negative Information

You may have noticed when you first looked at the images presented earlier in this chapter that you tended to like some of the people and to dislike others. It is not surprising that you had these emotions—these initial affective reactions are an essential and highly adaptive part of person perception. One of the things that we need to determine when we first perceive someone is whether that person poses any threat to our well-being. We may dislike or experience negative emotions about people because we feel that they are likely to harm us, just as we may like and feel positively about them if we feel that they can help us (Rozin & Royzman, 2001). Research has found that the threat and the trustworthiness of others are particularly quickly perceived, at least by people who are not trying to hide their intentions (Bar, Neta, & Linz, 2006; Todorov, Said, Engel, & Oosterhof, 2008).

Most people with whom we interact are not dangerous, nor do they create problems for us. In fact, when we are asked to rate how much we like complete strangers, we generally rate them positively (Sears, 1986). Because we generally expect people to be positive, people who are negative or threatening are salient, likely to create strong emotional responses, and relatively easy to spot.

Compared with positive information, negative information about a person tends to elicit more physiological arousal, draw greater attention, and exert greater impact on our judgments and impressions of the person. Hansen and Hansen (1988) had undergraduate students complete a series of trials in which they were shown, for very brief time periods, "crowds" of nine faces ([Figure 5.7, "Faces"](#)). On some of the trials, all the faces were happy or all the faces were angry. On other trials, the "crowd" was made up of eight happy faces and one angry face, or eight angry faces and one happy face. For each trial, the participants were instructed to say, as quickly as possible, whether the crowd contained a discrepant face or not. Hansen and Hansen found that the students were significantly faster at identifying the single angry face among the eight happy ones than they were at identifying the single happy face among the eight angry ones. They also made significantly fewer errors doing so. The researchers' con-

gry, rather than neutral, expressions, and Dijksterhuis and Aarts (2003) found that people co[u] more accurately recognize negative, rather than positive, words.
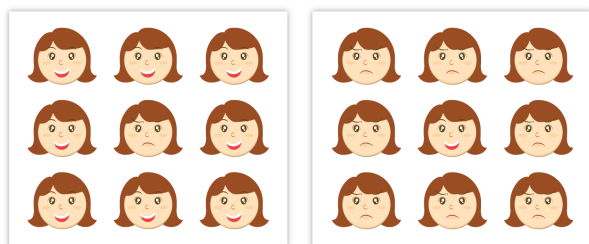

Figure 5.7 Faces.

Because negative faces are more salient and therefore more likely to grab our attention than are positive faces, people are faster at locating a single negative face in a display of positive faces than they are to locate a single positive face in a display of negative faces.

Our brains seem to be hardwired to detect negative behaviors (Adams, Gordon, Baird, Ambady, & Kleck, 2003), and at an evolutionary level this makes sense. It is important to tell the "good guys" from the "bad guys" and to try to avoid interacting with the latter. In one study, Tiffany Ito and her colleagues (Ito, Larsen, Smith, & Cacioppo, 1998) showed college students a series of positive, negative, and neutral images while their event-related brain potentials were collected. The researchers found that different parts of the brain reacted to positive and negative images and that the response to negative images was greater overall. They concluded that "negative information weighs more heavily on the brain" (p. 887). In sum, the results of research in person perception are clear: when we are perceiving people, negative information is simply more influential than positive information (Pratto & John, 1991).

Social Psychology in the Public Interest

Detecting Deception

One important person perception task that we must all engage in sometimes is to try to determine whether other people are lying to us. We might wonder whether our poker opponent is bluffing, whether our partner is being honest when she tells us she loves us, or whether our boss is really planning to give us the promotion he has promised. This task is particularly important for members of courtroom juries, who are asked determine the truth or falsehood of the testimony given by witnesses. And detecting deception is perhaps even more important for those whose job is to provide public security. How good are professionals, such as airport security officers and police detectives at determining whether or not someone is telling the truth?

It turns out that the average person is only moderately good at detecting deception and that experts do not seem to

lies and truths about 54% of the time (chance performance is 50%). This is not a big advanta[ge] [Increase Font Size] could have at least some practical consequences and that suggests that we can at least detect some deception. How-ever, the meta-analysis also found that experts—including police officers, detectives, judges, interrogators, crimi-nals, customs officials, mental health professionals, polygraph examiners, job interviewers, federal agents, and au-ditors—were not significantly better at detecting deception than were nonexperts.

Why is it so difficult for us to detect liars? One reason is that people do not expect to be lied to. Most people are good and honest folk, and we expect them to tell the truth, and we tend to give them the benefit of the doubt (Buller, Stiff, & Burgoon, 1996; Gilbert, Krull, & Malone, 1990). In fact, people are more likely to expect decep-tion when they view someone on a videotape than when they are having an interpersonal interaction with the per-son. It's as if we expect the people who are right around us to be truthful (Bond & DePaulo, 2006).

A second reason is that most people are pretty good liars. The cues that liars give off are quite faint, particularly when the lies that they are telling are not all that important. Bella DePaulo and her colleagues (DePaulo et al., 2003) found that in most cases it was very difficult to tell if someone was lying, although it was easier when the liar was trying to cover up something important (e.g., a sexual transgression) than when he or she was lying about something less important. De Paulo and colleagues did find, however, that there were some reliable cues to deception.

Compared with truth tellers, liars:

- Made more negative statements overall

- Appeared more tense

- Provided fewer details in their stories

- Gave accounts that were more indirect and less personal

- Took longer to respond to questions and exhibited more silent pauses when they were not able to prepare their responses

- Gave responses that were briefer and spoken in a higher pitch

A third reason it is difficult for us to detect liars is that we tend to think we are better at catching lies than we actu-ally are. This overconfidence may prevent us from working as hard as we should to try to uncover the truth.

Finally, most of us do not really have a very good idea of how to detect deception; we tend to pay attention to the wrong things. Many people think that a person who is lying will avert his or her gaze or will not smile or that per-haps he or she will smile too much. But it turns out that faces are not that revealing. The problem is that liars can more easily control their facial expressions than they can control other parts of their bodies. In fact, Ekman and Friesen (1974) found that people were better able to detect other people's true emotions when they could see their bodies but not their faces than when they could see their faces but not their bodies. Although we may think that de-ceivers do not smile when they are lying, it is actually common for them to mask their statements with false smiles

Previous: 5. Perceiving Others

Next: Inferring Dispositions Using Causal Attribution

Recently, advances in technology have begun to provide new ways to assess deception. Som̄ Increase Font Size the language of truth tellers, other software analyzes facial microexpressions that are linked with lying (Newman, Pennebaker, Berry, & Richards, 2003), and still other software uses neuroimaging techniques to try to catch liars (Langleben et al., 2005). Whether these techniques will be successful, however, remains to be seen.

Judging People by Their Traits

Although we can learn some things about others by observing their physical characteristics and their nonverbal behaviors, to really understand them we will eventually need to know their personality traits. Traits are important because they are the basic language by which we understand and communicate about people. When we talk about other people, we describe their traits. Our friends are "fun," "creative," and "crazy in a good way," or "quiet," "serious," and "controlling." The language of traits is a powerful one—indeed, there are over 18,000 trait terms in the English language.

Combining Traits: Information Integration

Let's consider for a moment how people might use trait terms to form an overall evaluation of another person. Imagine that you have to describe two friends of yours, Amir and Connor, to another person, Rianna, who might be interested in dating one of them. You'll probably describe the two men in terms of their physical features first, but then you'll want to say something about their personalities. Let's say that you want to make both Amir and Connor sound as good as possible to Rianna, but you also want to be honest and not influence her one way or the other. How would you do that? You would probably start by mentioning their positive traits: Amir is "intelligent" and "serious"; Connor is "fun" and "exciting." But to be fair, you would also need to mention their negative traits: Amir sometimes seems "depressed," and Connor can be "inconsiderate."

You might figure that Rianna will just combine whatever information you give her, perhaps in a mathematical way. For instance, she might listen to all the traits that you mention, decide how positive or negative each one is, and then add the traits together or average them. Research has found that people do exactly that, both for strangers and for people whom they know very well (Anderson, 1974; Falconi & Mullet, 2003). Consider what might happen if you gave Rianna the following information:

- Amir is *smart, serious, kind, and sad*.

- Connor is *fun, happy, selfish, and inconsiderate*.

Rianna might decide to score each trait on a scale of +5 (very positive) to –5 (very negative). Once she has these numbers, she could then either add them together or average them to get an overall judgment.

Amir	
Smart	+5
Serious	+1
Kind	+4
Sad	–4
Sum	**+6.0**
Average	+1.5
Connor	
Fun	+3
Happy	+2
Selfish	–4
Inconsiderate	–5
Sum	**–4.0**
Average	–1.0

Based on this scoring, Rianna would probably decide that she likes Amir more than Connor. Of course, different people might weight the traits in somewhat different ways, and this would lead different people to draw different impressions about Amir and Connor. But there is pretty good agreement among most people about the meaning of traits, at least in terms of the overall positivity or negativity of each trait, and thus most people would be likely to draw similar conclusions.

Now imagine that you later thought of some other new, moderately positive characteristics about Amir—that he was also "careful" and "helpful." Whether you told Rianna about them might depend on how you thought they would affect her overall impression of Amir. Perhaps these new traits would make Rianna like Amir more (after all, they do add new positive information about him). But perhaps they might make her like him less (if the new, moderately positive information diluted the existing positive impression she has already formed about him).

One way to think about this is to consider whether Rianna might be *adding* the traits together or *averaging* them. In our first example, it didn't matter because the outcome was the same. But now it might—if she's adding the traits together, then Rianna will probably like Amir more after she hears the new information, because new positive traits have been added to the existing sum score. If she is averaging the traits together, however, then Rianna will probably like him less than she did before, because the new, more moderate information tends to dilute the initial impressions.

It turns out that in most cases, our judgments are better predicted by mental averaging than by mental adding (Mills, 2007). What this means is that when you are telling someone about another person and you are trying to get him or her to like the person, you should say the most positive things that you know but leave out the more moder-

The Importance of the Central Traits Warm and Cold

Although the averaging model is quite good at predicting final impressions, it is not perfect. This is because some traits are simply weighted more heavily than others. For one, negative information is more heavily weighted than is positive information (Rozin & Royzman, 2001). In addition to the heavy weight that we give to negative traits, we give a particular emphasis to the traits "warm" and "cold." Imagine two men, Brad and Phil, who were described with these two sets of characteristics:

- Brad is *industrious, critical, warm, practical, and determined*.

- Phil is *industrious, critical, cold, practical, and determined*.

As you can see, the descriptions are identical except for the presence of "warm" and "cold." In a classic study, Solomon Asch (1946) found that people described with these two sets of traits were perceived very differently— the "warm" person very positively and the "cold" person very negatively.

To test whether or not these differences would influence real behavior, Harold Kelley (1950) had students read about a professor who was described either as "rather cold" or as "very warm." Then the professor came into the classroom and led a 20-minute discussion group with the students. Although the professor behaved in the same way for both groups, the students nevertheless reacted very differently to him. The students who were expecting the "warm" instructor were more likely to participate in the discussion, in comparison with those who were expecting him to be "cold." And at the end of the discussion, the students also rated the professor who had been described as "warm" as being significantly more humorous, sociable, popular, and better natured than the "cold" professor. Moreover, the effects of warmth and coolness seem to be wired into our bodily responses. Research has found that even holding a cup of hot versus iced coffee, or making judgments in warm versus cold rooms leads people to judge others more positively (Ijzerman & Semin, 2009; Williams & Bargh, 2008).

In short, the particular dimension *warm* versus *cold* makes a big difference in how we perceive people—much bigger than do other traits. As a result, the traits of warm and cold are known as **central traits,** which are *characteristics that have a very strong influence on our impressions of others* (Asch, 1946). The powerful influence of central traits is due to two things. One, they lead us to make inferences about other traits that might not have been mentioned. The students who heard that the professor was "warm" might also have assumed that he had other positive traits (maybe "nice" and "funny"), in comparison with those who heard that he was "cold." Two, the important central traits also color our perceptions of the other traits that surround them. When a person is described as "warm" and "intelligent," the meaning of "intelligent" seems a lot better than does the term "intelligent" in the context of a person who is also "cold." Overall, the message is clear: if you want to get someone to like you, try to act in a warm manner toward them. Be friendly, nice, and interested in what they say. This attention you pay to the other will be more powerful than any other characteristics that you might try to display to them. The importance of perceptions of warmth-coldness has been confirmed in many other contexts. For example, in the field of psychotherapy, many studies have indicated that therapists' warmth, empathy, and genuineness are the three most important traits in establishing a strong and trusting relationship with clients, which in turn leads to positive change

First Impressions Matter: The Primacy Effect

It has frequently been said that "first impressions matter." Social psychological research supports this idea. The **primacy effect** describes *the tendency for information that we learn first to be weighted more heavily than is information that we learn later*. One demonstration of the primacy effect was conducted by Solomon Asch (1946). In his research, participants learned some traits about a person and then made judgments about him. One half of the participants saw this list of traits:

- *Intelligent, industrious, impulsive, critical, stubborn, envious*

The other half of the participants saw this list:

- *Envious, stubborn, critical, impulsive, industrious, intelligent*

You may have noticed something interesting about these two lists—they contain exactly the same traits but in reverse order.

Asch discovered something interesting in his study: because the traits were the same, we might have expected that both groups would form the same impression of the person, but this was not at all the case. Rather, Asch found that the participants who heard the first list, in which the positive traits came first, formed much more favorable impressions than did those who heard the second list, in which the negative traits came first. Similar findings were found by Edward Jones (1968), who had participants watch one of two videotapes of a woman taking an intelligence test. In each video, the woman correctly answered the same number of questions and got the same number wrong. However, when the woman got most of her correct answers in the beginning of the test but got more wrong near the end, she was seen as more intelligent than when she got the same number correct but got more correct at the end of the test.

Primacy effects also show up in other domains, even in those that seem really important. For instance, Koppell and Steen (2004) found that in elections in New York City, the candidate who was listed first on the ballot was elected more than 70% of the time, and Miller and Krosnick (1998) found similar effects for candidate preferences in laboratory studies.

This is not to say that it is always good to be first. In some cases, the information that comes last can be most influential. **Recency effects**, *in which information that comes later is given more weight*, although much less common than primacy effects, may sometimes occur. For example, de Bruin (2005) found that in competitions such as the Eurovision Song Contest and ice skating, higher marks were given to competitors who performed last.

Considering the primacy effect in terms of the cognitive processes central to human information processing leads us to understand why it can be so powerful. One reason is that humans are cognitive misers. Because we desire to conserve our energy, we are more likely to pay more attention to the information that comes first and less likely to attend to information that comes later. In fact, when people read a series of statements about a person, the amount

wide awake and when we are distracted than when we are paying attention (Webster, Richter, 1996).

Another reason for the primacy effect is that the early traits lead us to form an initial expectancy about the person, and once that expectancy is formed, we tend to process information in ways that keep that expectancy intact. Thinking back to Chapter 2 and the discussion of social cognition, we can see that this of course is a classic case of assimilation—once we have developed a schema, it becomes difficult to change it. If we learn that a person is "intelligent" and "industrious," those traits become cognitively accessible, which leads us to develop a positive expectancy about the person. When the information about the negative features comes later, these negatives will be assimilated into the existing knowledge more than the existing knowledge is accommodated to fit the new information. Once we have formed a positive impression, the new negative information just doesn't seem as bad as it might have been had we learned it first. This is an important factor in explaining the **halo effect**, which is *the influence of a global positive evaluation of a person on perceptions of their specific traits*. Put simply, if we get an initially positive general impression of someone, we often see their specific traits more positively. The halo effect has been demonstrated in many social contexts, including a classic investigation by Bingham and Moore (1931) on job interviewing and a far more recent study of students' evaluations of their professors (Keeley, English, Irons, & Hensley, 2013).

You can be sure that it would be good to take advantage of the primacy and halo effects if you are trying to get someone you just met to like you. Begin with your positive characteristics, and only bring the negatives up later. This will create a much better outcome than beginning with the negatives.

Key Takeaways

- Every day we must size up the people we interact with. The process of doing this is known as person perception.

- We can form a wide variety of initial impressions of others quickly and often quite accurately.

- Nonverbal behavior is communication that does not involve speaking, including facial expressions, body language, touching, voice patterns, and interpersonal distance. We rely on nonverbal behavior in our initial judgments of others.

- The particular nonverbal behaviors that we use, as well as their meanings, are determined by social norms, and these may vary across cultures.

- In comparison with positive information about people, negative information tends to elicit more physiological arousal, draw greater attention, and exert greater impact on our judgments and impressions of people.

- People are only moderately good at detecting deception, and experts are not usually much better than the average person.

- The primacy effect occurs because we pay more attention to information that come [Increase Font Size] cause initial information colors how we perceive information that comes later.
- These processes also help to explain how the halo effect occurs.

Exercises and Critical Thinking

1. Consider a case where you formed an impression of someone quickly and on only a little information. How accurate do you think your judgment was and why? What information did you take into account? What information might you have missed?

2. Consider some of the nonverbal behaviors that you and your friends use when you communicate. What information are you usually trying to communicate by using them? When do you find yourself using more vigorous gesturing and why?

3. Give an example of a situation in which you have noticed the effects of central traits on your perception of someone. Why do you think that this happened?

4. Describe a situation where you were influenced by either the primacy or the halo effect in your initial perceptions of someone. How accurate did those initial perceptions turn out to be and why?

References

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., Schaller, M. (2006). They all look the same to me (unless they're angry): From out-group homogeneity to out-group heterogeneity. *Psychological Science, 17*(10), 836–840.

Adams, R. B., Jr., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science, 300*(5625), 1536.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 201–271). San Diego, CA: Academic Press.

Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology, 16*(1), 4–13. doi:

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from nonverbal behavior and physical attractiveness. *Journal of Personality and Social Psychology, 64(3)*, 431–441.

Anderson, N. H. (1974). Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1–101). New York, NY: Academic Press;

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258–290.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269–278. doi: 10.1037/1528–3542.6.2.269;

Belmore, S. M., & Hubbard, M. L. (1987). The role of advance expectancies in person memory. *Journal of Personality and Social Psychology, 53*(1), 61–70.

Bingham, W. V., & Moore, B. V. (1931). *How to interview.* Oxford, England: Harpers.

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234.

Buller, D. B., Stiff, J. B., & Burgoon, J. K. (1996). Behavioral adaptation in deceptive transactions: Fact or fiction: Reply to Levine and McCornack. *Human Communication Research*, *22*(4), 589-603. doi:10.1111/j.1468-2958.1996.tb00381.x

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology, 89*(6), 884–898;

Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception, 34*(10), 1171–1180;

de Bruin, W. B. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychologica, 118*(3), 245–260. doi: 10.1016/j.actpsy.2004.08.005

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118.

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science, 14*(1), 14–18.

Ekman, P., & Davidson, R. J. (1993). Voluntary smiling changes regional brain activity. *Psychological*

Previous: 5. Perceiving Others

Next: Inferring Dispositions Using Causal Attribution

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal* [Increase Font Size] *and Social Psychology, 29*(3), 288–298. doi: 10.1037/h0036006

Falconi, A., & Mullet, E. (2003). Cognitive algebra of love through the adult life. *International Journal of Aging and Human Development, 57*(3), 275–290.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*(4), 571–583.

Frank, M. G., & Ekman, P. (1993). Not all smiles are created equal: The differences between enjoyment and nonenjoyment smiles. *Humor: International Journal of Humor Research, 6*(1), 9–26.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology, 59*(4), 601–613.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917–924.

Haselton, M. G., & Funder, D. C. (2006). The evolution of accuracy and bias in social judgment. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 15–37). Madison, CT: Psychosocial Press.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–233.

Heberlein, A. S., Adolphs, R., Tranel, D., & Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience, 16*(7), 1143–1158.

Hood, B. M., & Macrae, C. N. (2007). Look into my eyes: The effect of direct gaze on face processing in children and adults. In R. Flom, K. Lee, & D. Muir (Eds.), *Gaze-following: Its development and significance* (pp. 283–296). Mahwah, NJ: Lawrence Erlbaum; Mason, M. F.,

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137* (2), 297–315.

Ijzerman, H., & Semin, G. R. (2009). The thermometer of social relations: Mapping social proximity on temperature. *Psychological Science, 20*(10), 1214–1220

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75*(4), 887–900.

Johnson, K. L., Gill, S., Reichman, V., & Tassinary, L. G. (2007). Swagger, sway, and sexuality: Judg-

Jones, E. E. (1968). Pattern of performance and ability attribution: An unexpected prin[Increase Font Size] effect. *Journal of Personality and Social Psychology, 10*(4), 317–340.

Keeley, J. W., English, T., Irons, J., & Henslee, A. M. (2013). Investigating halo and ceiling effects in student evaluations of instruction. *Educational And Psychological Measurement*, *73*(3), 440-457. doi:10.1177/0013164412475300

Kelley, H. H. (1950). The warm-cold variable in first impressions of persons. *Journal of Personality,18*(4), 431–439.

Knapp, M. L., & Hall, J. A. (2006). *Nonverbal communication in human interaction* (6th ed.). Belmont, CA: Thomson Wadsworth.

Koppell, J. G. S., & Steen, J. A. (2004). The effects of ballot position on election outcomes. *Journal of Politics, 66*(1), 267–281.

Krauss, R. M., Chen, Y., & Chawla, P. (Eds.). (1996). *Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us?* San Diego, CA: Academic Press.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human Brain Mapping, 26*(4), 262–272.

Macrae, C. N., & Quadflieg, S. (2010). Perceiving people. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 428–463). Hoboken, NJ: John Wiley & Sons.

Manstead, A. S. R. (Ed.). (1991). *Expressiveness as an individual difference*. New York, NY: Cambridge University Press;

Mason, M. F., & Macrae, C. N. (2004). Categorizing and individuating others: The neural substrates of person perception. *Journal of Cognitive Neuroscience, 16*(10), 1785–1795. doi: 10.1162/0898929042947801

Miller, J. M., & Krosnick, J. A. (1998). The impact of candidate name order on election outcomes. *Public Opinion Quarterly, 62*(3), 291–330.

Mills, J. (2007). Evidence forming attitudes from combining beliefs about positive attributes of activities follows averaging (Unpublished manuscript). University of Maryland, College Park.

Montepare, J. M., & Zebrowitz-McArthur, L. (1988). Impressions of people created by age-related qualities of their gaits. *Journal of Personality and Social Psychology, 55*(4), 547–556.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting

Pennebaker, J. W., Rime, B., & Blankenship, V. E. (1996). Stereotypes of emotional ex Northerners and Southerners: A cross-cultural test of Montesquieu's hypotheses. *Journal of Personality and Social Psychology, 70*(2), 372–380.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*(3), 380–391.

Richeson, J. A., & Shelton, J. N. (2005). Brief report: Thin slices of racial bias. *Journal of Nonverbal Behavior, 29*(1), 75–86.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5*(4), 296–320.

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PLoS ONE, 5*(1), e8733;

Rule, N. O., Ambady, N., & Hallett, K. C. (2009). Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features. *Journal of Experimental Social Psychology, 45*(6), 1245–1251.

Rule, N. O., Ambady, N., Adams, R. B., Jr., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology, 95*(5), 1019–1028;

Rule, N. O., Ishii, K., Ambady, N., Rosen, K. S., & Hallett, K. C. (2011). Found in translation: Cross-cultural consensus in the accurate categorization of male sexual orientation. *Personality And Social Psychology Bulletin*, *37*(11), 1499-1507. doi:10.1177/0146167211415630

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*(3), 515–530.

Shapiro, D. A. (1969). Empathy, warmth, and genuineness in psychotherapy. *British Journal of Clinical and Social Psychology, 8(4),* 350-361. doi: 10.1111/j.2044-8260.1969.tb00627.x

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science, 308*(5728), 1623–1626.

Todorov, A., Said, C. P., Engel, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*(12), 455–460. doi: 10.1016/j.tics.2008.10.001

Turati, C., Cassia, V. M., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Development, 77*(2), 297–311.

Walker-Andrews, A. S. (2008). Intermodal emotional processes in infancy. In M. Lewis, J. M. Havi-

Webster, D. M., Richter, L., & Kruglanski, A. W. (1996). On leaping to conclusions wh Mental fatigue effects on impressional primacy. *Journal of Experimental Social Psychology, 32*(2), 181–195.

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science, 322*(5901), 606–607.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100ms exposure to a face. *Psychological Science, 17*(7), 592–598.

Increase Font Size

LICENSE

SHARE THIS BOOK

Powered by Pressbooks

Guides and Tutorials

Increase Font Size

Previous: 5. Perceiving Others

Next: Inferring Dispositions Using Causal Attribution

First Impressions Matter: A Model of Confirmatory Bias
Author(s): Matthew Rabin and Joel L. Schrag
Source: *The Quarterly Journal of Economics,* Vol. 114, No. 1 (Feb., 1999), pp. 37–82
Published by: Oxford University Press
Stable URL: https://www.jstor.org/stable/2586947
Accessed: 11-03-2020 02:36 UTC

FIRST IMPRESSIONS MATTER:
A MODEL OF CONFIRMATORY BIAS*

MATTHEW RABIN AND JOEL L. SCHRAG

Psychological research indicates that people have a cognitive bias that leads them to misinterpret new information as supporting previously held hypotheses. We show in a simple model that such *confirmatory bias* induces overconfidence: given any probabilistic assessment by an agent that one of two hypotheses is true, the appropriate beliefs would deem it less likely to be true. Indeed, the hypothesis that the agent believes in may be more likely to be *wrong* than right. We also show that the agent may come to believe with near certainty in a false hypothesis despite receiving an infinite amount of information.

> The human understanding when it has once adopted an opinion draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusion may remain inviolate.
>
> Francis Bacon[1]

I. INTRODUCTION

How do people form beliefs in situations of uncertainty? Economists have traditionally assumed that people begin with subjective beliefs over the different possible states of the world and use Bayes' Rule to update those beliefs. This elegant and powerful model of economic agents as Bayesian statisticians is the foundation of modern information economics.

Yet a large and growing body of psychological research

1. From *The New Organon and Related Writings* [1960; 1620], quoted in Nisbett and Ross [1980, p. 167].

suggests that the way people process information often departs systematically from Bayesian updating. In this paper we formally model and explore the consequences of one particular departure from Bayesian rationality: *confirmatory bias.* A person suffers from confirmatory bias if he tends to misinterpret ambiguous evidence as confirming his current hypotheses about the world. Teachers misread performance of pupils as supporting their initial impressions of those pupils; many people misread their observations of individual behavior as supporting their prior stereotypes about groups to which these individuals belong; scientists biasedly interpret data as supporting their hypotheses.

Our simple model by and large confirms an intuition common in the psychology literature: confirmatory bias leads to overconfidence, in the sense that people on average believe more strongly than they should in their favored hypotheses. The model also yields surprising further results. An agent who suffers from confirmatory bias may come to believe in a hypothesis that is *probably wrong,* meaning that a Bayesian observer who was aware of the agent's confirmatory bias would, after observing the agent's beliefs, favor a different hypothesis than the agent. We also show that even an infinite amount of information does not necessarily overcome the effects of confirmatory bias: over time an agent may with positive probability come to believe with near certainty in the wrong hypothesis.

In Section II—which readers impatient for math may wish to skip—we review some of the psychological evidence that humans are prone to confirmatory bias. In Section III we present our formal model and provide examples and general propositions illustrating the implications of confirmatory bias. In our model, an agent initially believes that each of two possible states of the world is equally likely. The agent then receives a series of independent and identically distributed signals that are correlated with the true state. To model confirmatory bias, we assume that when the agent gets a signal that is counter to the hypothesis he currently believes is more likely, there is a positive probability that he misreads that signal as *supporting* his current hypothesis. The agent is unaware that he is misreading evidence in this way and engages in Bayesian updating that would be fully rational given his environment if he were not misreading evidence.[2]

2. Researchers have, of course, documented many other biases in information processing. We develop a model ignoring these other biases, assuming complete rationality except for this one bias, so as to keep our model tractable and because we feel that incorporating documented biases into the Bayesian model one at a time is useful for carefully identifying the effects of each particular bias.

Because we assume that the agent always correctly interprets evidence that confirms his current beliefs, relative to *proper* Bayesian updating he is biased toward confirming his current hypothesis.

So, for example, a teacher may believe either that Marta is smarter than Bart or that Bart is smarter than Marta; he initially believes each is equally likely, and over time he collects a series of signals that help him to identify who is smarter. If, after receiving one or more signals, the teacher believes that Marta is probably smarter than Bart, confirmatory bias may lead him to erroneously interpret his next signal as supporting this hypothesis. Therefore, the teacher's updated belief that Marta is smarter than Bart may be stronger than is warranted.

The notion that the teacher is likely to believe "too strongly" that Marta is smarter corresponds to the commonly held intuition that confirmatory bias leads to overconfidence. While qualifying this intuition with several caveats, our model by and large confirms it: given any probabilistic assessment by an agent that one of the hypotheses is probably true, the appropriate beliefs should on average deem it less likely to be true. Intuitively, a person who believes strongly in a hypothesis is likely to have misinterpreted some signals that conflict with what he believes, and hence is likely to have received more evidence against his believed hypothesis than he realizes.

Our analysis shows that a more surprising result arises when confirmatory bias is severe: a Bayesian observer with no direct information of her own, but who can observe the agent's belief in favor of one hypothesis, may herself believe that the *other* hypothesis is more likely. We show that such "wrongness" can arise when the agent's evidence is sufficiently mixed. Intuitively, if the agent has perceived almost as much evidence against his hypothesis as supporting it, then, since some of the evidence he perceives as supportive is actually *not* supportive, it is likely that a majority of the real signals oppose his hypothesis. Because such wrongness only arises when the agent has relatively weak evidence supporting his favored hypothesis, however, the agent *on average* correctly judges which of the two hypotheses is more likely, in the sense that his best guess is right most of the time.

While seemingly straightforward, the intuition for our overconfidence and wrongness results conceals some subtle implications of the agent's confirmatory bias. For example, an agent who currently believes in Hypothesis A (say) may *once* have believed in Hypothesis B, at which time he had a propensity to misread as

supporting Hypothesis B evidence actually in favor of Hypothesis A. But then the agent may underestimate how many signals supporting Hypothesis A he has received, and thus he may be *under*confident in his belief in favor of Hypothesis A. Indeed, we show that an agent who has only recently come to believe in a hypothesis is likely to be underconfident in that hypothesis, because until recently he has been biased against his current hypothesis. If a teacher used to think Bart was smarter than Marta and only recently concluded that Marta is smarter, then probably he has been ignoring evidence all along that Marta is smarter. The simple overconfidence and wrongness results hold because an agent has probably believed in his currently held hypothesis during most of the time he has been receiving information and so, *on average,* has been biased toward this hypothesis.

In Section IV we investigate the implications of confirmatory bias after the agent receives an infinite sequence of signals. In the absence of confirmatory bias, an agent will always come to believe with near certainty in the correct hypothesis if he receives an infinite sequence of signals. If the confirmatory bias is sufficiently severe or the strength of individual signals is weak, however, then with positive probability the agent may come to believe with near certainty that the *incorrect* hypothesis is true. Intuitively, once the agent comes to believe in an incorrect hypothesis, the confirmatory bias inhibits his ability to overturn his erroneous beliefs. If the bias is strong enough, the expected drift once the agent comes to believe in the false hypothesis is toward believing more strongly in that hypothesis, guaranteeing a positive probability that the agent ends up forever believing very strongly in the false hypothesis. The results of Section IV belie the common intuition that learning will *eventually* correct cognitive biases. While this is true for sufficiently mild confirmatory bias, when the bias is sufficiently severe "learning" can *exacerbate* the bias.

The premise of this paper is that explicit formalizations of departures from Bayesian information processing are crucial to incorporating psychological biases into economic analysis. For the most part, we do not in this paper take the important next step of developing extended economic applications of the bias we model. In Section V, however, we illustrate one implication of confirmatory bias by sketching a simple principal-agent model. We illustrate how a principal may wish to mute the incentives that she offers an agent who suffers from confirmatory bias. Indeed, we

show that even if it is very easy for an agent to gather information, so that a principal can at negligible costs provide incentives for an agent to search for profitable investment opportunities, the principal may choose *not* to provide these incentives to a confirmatory agent. This arises when the expected costs in terms of an overconfident agent investing too much in risky projects outweigh the expected benefit of the agent being better informed. We conclude in Section VI by discussing some other potential economic implications of confirmatory bias, as well as highlighting some likely obstacles to applying our model.

II. A Review of the Psychology Literature

Many different strands of psychological research yield evidence on phenomena that we are modeling under the rubric of confirmatory bias. Before reviewing this literature, we first wish to distinguish a form of "quasi-Bayesian" information processing from the bias we are examining. Although the two phenomena are related—and not always distinguished clearly in the psychology literature—they differ importantly in their implications for decision theory. Suppose that, once they form a strong hypothesis, people simply stop being attentive to relevant new information that contradicts or supports their hypotheses. Intuitively, when you become convinced that one investment strategy is more lucrative than another, you may simply stop paying attention to even freely available additional information.[3]

Bruner and Potter [1964] elegantly demonstrate such anchoring. About 90 subjects were shown blurred pictures that were gradually brought into sharper focus. Different subjects began viewing the pictures at different points in the focusing process, but the pace of the focusing process and final degree of focus were identical for all subjects. Strikingly, of those subjects who began their viewing at a severe-blur stage, less than a quarter eventually identified the pictures correctly, whereas over half of those who began viewing at a light-blur stage were able to correctly identify the pictures. Bruner and Potter [p. 424] conclude that "Interference may be accounted for partly by the difficulty of rejecting incorrect hypotheses based on substandard cues." That

---

3. Such behavior corresponds to a natural economic "cognitive-search" model: if we posit a cost to information processing, in many settings the natural stopping rule would be to process information until beliefs are sufficiently strong in one direction or another, and then stop.

is, people who use weak evidence to form initial hypotheses have difficulty correctly interpreting subsequent, better information that contradicts those initial hypotheses.[4]

This form of anchoring does not necessarily imply that people *misinterpret* additional evidence to either disconfirm *or* confirm initial hypotheses, only that they ignore additional evidence. Such a tendency to anchor on initial hypotheses can therefore be reconciled with Bayesian information processing. While such anchoring is potentially quite important, psychological evidence reveals a stronger and more provocative phenomenon: people tend to *misread* evidence as additional support for initial hypotheses. If a teacher initially believes that one student is smarter than another, she has the propensity to confirm that hypothesis when interpreting later performance.[5] Lord, Ross, and Lepper [1979, p. 2099] posited some of the underlying cognitive mechanisms involved in such propensities:

> . . . there is considerable evidence that people tend to interpret subsequent evidence so as to maintain their initial beliefs. The biased assimilation processes underlying this effect may include a propensity to remember the strengths of confirming evidence but the weaknesses of disconfirming evidence, to judge confirming evidence as relevant and reliable but disconfirming evidence as irrelevant and unreliable, and to accept confirming evidence at face value while scrutinizing disconfirming evidence hypercritically. With confirming evidence, we suspect that both lay and professional scientists rapidly reduce the complexity of the information and remember only a few well-chosen supportive impressions. With disconfirming evidence, they continue to reflect upon any information that suggests less damaging "alternative interpretations." Indeed, they may even come to regard the ambiguities and conceptual flaws in the data *opposing* their hypotheses as somehow suggestive of the fundamental *correctness* of those hypotheses. Thus, com-

---

4. A similar experiment [Wyatt and Campbell 1951] was cited by Perkins [1981] as one interpretation of the perspective that "fresh" thinkers may be better at seeing solutions to problems than people who have meditated at length on the problems, because the fresh thinkers are not overwhelmed by the "interference" of old hypotheses.

5. A related arena where the confirmation bias has been studied widely is in counselor judgments: counselors in clinical settings tend to confirm original suppositions in their eventual judgments. If you are told ahead of time that an interviewee is combative, then both your conduct and your interpretation of his conduct during an interview may reinforce that supposition, even if he is in fact no more combative than the average person. See, e.g., Haverkamp [1993]. There has also been extensive research on confirmatory bias in the interviewing process more generally; see, e.g., Dougherty, Turban, and Callender [1994] and Macan and Dipboye [1994]. Research applying variants of confirmatory bias to other domains includes Arkes [1989] and Borum, Otto, and Golding [1993] to the law; Baumann, Deber, and Thompson [1991] to medicine; and Souter [1993] discusses the implications of overconfidence to business insurance.

pletely inconsistent or even *random* data—when "processed" in a suitably biased fashion—can maintain or even reinforce one's preconceptions.

The most striking evidence for the confirmatory bias is a series of experiments demonstrating how providing the *same* ambiguous information to people who differ in their initial beliefs on some topic can move their beliefs *farther apart.* To illustrate such polarization, Lord, Ross, and Lepper [1979] asked 151 undergraduates to complete a questionnaire that included three questions on capital punishment. Later, 48 of these students were recruited to participate in another experiment. Twenty-four of them were selected because their answers to the earlier question-naire indicated that they were " 'proponents' who favored capital punishment, believed it to have a deterrent effect, and thought most of the relevant research supported their own beliefs. Twenty-four were opponents who opposed capital punishment, doubted its deterrent effect and thought that the relevant research supported *their* views." These subjects were then asked to judge the merits of randomly selected studies on the deterrent efficacy of the death penalty, and to state whether a given study (along with criticisms of that study) provided evidence for or against the deterrence hypothesis. Subjects were then asked to *rate,* on 16 point scales ranging from -8 to $+8$, how the studies they had read moved their attitudes toward the death penalty, and how they had changed their beliefs regarding its deterrent efficacy. Lord, Ross, and Lepper [pp. 2102–2104] summarize the basic results (all of which hold with confidence $p < .01$) as follows:

> The relevant data provide strong support for the polarization hypothe-sis. Asked for their final attitudes relative to the experiment's start, proponents reported that they were *more* in favor of capital punishment, whereas opponents reported that they were *less* in favor of capital punish-ment. . . . Similar results characterized subjects' beliefs about deterrent efficacy. Proponents reported greater belief in the deterrent effect of capital punishment, whereas opponents reported less belief in this deterrent effect.

Plous [1991] replicates the Lord-Ross-Lepper results in the context of judgments about the safety of nuclear technology. Pro- and antinuclear subjects were given identical information and arguments regarding the Three Mile Island nuclear disaster and a case of false military alert that could have led to the launching of U. S. nuclear missiles. Plous [p. 1068] found that 54 percent of pronuclear subjects became more pronuclear from the informa-tion, while only 7 percent became less pronuclear. By contrast,

only 7 percent of the antinuclear subjects became less antinuclear from the information while 45 percent became more antinuclear.[6]

Darley and Gross [1983] demonstrate a related and similarly striking form of polarization due to confirmatory bias. Seventy undergraduates were asked to assess a nine-year-old girl's academic skills in several different academic areas. Before completing this task, the students received information about the girl and her family and viewed a video tape of the girl playing in a playground. One group of subjects was given a fact sheet that described the girl's parents as college graduates who held white-collar jobs; these students viewed a video of the girl playing in what appeared to be a well-to-do, middle class neighborhood. The other group of subjects was given a fact sheet that described the girl's parents as high school graduates who held blue-collar jobs; these students viewed a video of the same girl playing in what appeared to be an impoverished inner-city neighborhood. Half of each group of subjects were then asked to evaluate the girl's reading level, measured in terms of equivalent grade level.[7] There was a small difference in the two groups' estimates—those subjects who had viewed the "inner-city" video rated the girl's skill level at an average of 3.90 (i.e., $9/10$ through third grade) while those who had viewed the "suburban video" rated the girl's skill level at an average of 4.29. The remaining subjects in each group were shown a second video of the girl answering (with mixed success) a series of questions. Afterwards, they were asked to

---

6. These percentages were derived from Table 2 of Plous [1991, p. 1068], aggregating across two studies; the remaining subjects in each case reported no change in beliefs. For other papers following on Lord, Ross, and Lepper [1979], see Fleming and Arrowood [1979]; Jennings, Lepper, and Ross [1981]; Hubbard [1984]; Lepper, Ross, and Lau [1986]. See also Miller, McHoskey, Bane, and Dowd [1993] for more mixed evidence regarding the Lord-Ross-Lepper experiment. In the passage above, Lord, Ross, and Lepper posit that even professional scientists are susceptible to such same-evidence polarization. Indeed, many economists and other academics have probably observed how differing schools of thought interpret ambiguous evidence differently. An example was once told to one of us by a colleague. He saw the same model—calibrating the elasticity of demand facing a Cournot oligopolist as a function of the number of firms in an industry—described at the University of Chicago and at the Massachusetts Institute of Technology. A Chicago economist derived the formula and said, "Look at how few firms you need to get close to infinite elasticities and perfect competition." An M.I.T. economist derived the same formula and said, "Look at how large n [the number of firms] has to be before you get anywhere close to an infinite elasticity and perfect competition." These different schools each interpreted the same *mathematical formula* as evidence reinforcing their respective views. For related analysis in the scientific domain, see also Mahoney [1977].

7. The subjects were also asked to evaluate the girl's mathematics and liberal arts skill levels; we report the results that are least supportive of the existence of confirmatory bias.

evaluate the girl's reading level. The inner-city video group rated the girl's skill level at an average of 3.71, significantly *below* the 3.90 estimate of the inner-city subjects who did not view the question-answer video. Meanwhile, the suburban video group rated the girl's skill level at an average of 4.67, significantly *above* the 4.29 estimate of the suburban subjects who did not view the second video. Even though the two groups viewed the *identical* question-and-answer video, the additional information further polarized their assessments of the girl's skill level. Darley and Gross interpret this result as evidence of confirmatory bias— subjects were influenced by the girl's background in their initial judgments, but their beliefs were evidently influenced even more strongly by the effect their initial hypotheses had on their interpretation of further evidence.[8]

Our reading of the psychology literature leads us to conclude that any of three different information-processing problems contribute to confirmatory bias. First, researchers widely recognize that confirmatory bias and overconfidence arise when people must interpret *ambiguous* evidence (see, e.g., Keren [1987] and Griffin and Tversky [1992]). Lord, Ross, and Lepper's [1979] study, discussed above, clearly illustrates the point. Keren [1988] notes the lack of confirmatory bias in visual perceptions and concludes that confirmatory tendency depends on some degree of abstraction and "discrimination" (i.e., the need for interpretation) not present in simple visual tasks. A primary mechanism of stereotype-maintenance is our tendency to interpret ambiguous behavior according to previous stereotype.[9] Similarly, a teacher may interpret an ambiguous answer by a student as either creative or just plain stupid, according to his earlier impressions of the student,

---

8. It should be noted that polarization of the form identified by Darley and Gross [1983] provides more direct evidence of confirmatory bias than does polarization identified by Lord, Ross, and Lepper [1979] and related papers. As Jeff Ely pointed out to us, Lord, Ross, and Lepper permit an alternative interpretation: that some people are predisposed to interpret ambiguous evidence one way and some the other. Hence, observing further polarization by groups who already differ may not reflect confirmatory bias per se, but underlying differences in interpretation of evidence that would appear irrespective of subjects' current beliefs. While this interpretation also departs from common-priors Bayesian information processing and will often yield similar implications as confirmatory bias, it is conceptually distinct and would sometimes yield different predictions. By demonstrating polarization based on differing beliefs induced in two ex ante identical groups of subjects, Darley and Gross are not subject to this alternative interpretation.

9. A vast literature explores the mechanisms by which people retain ethnic, gender, and other group stereotypes. See, e.g., Hamilton and Rose [1980]; Bodenhausen and Wyer [1985]; Bodenhausen and Lichtenstein [1987]; Stangor [1988]; Stangor and Ruble [1989]; and Hamilton, Sherman, and Ruvolo [1990].

but will be less likely to biasedly interpret more objective feedback such as answers to multiple-choice questions.

Second, confirmatory bias can arise when people must interpret statistical evidence to assess the correlation between phenomena that are separated by time. Nisbett and Ross [1980] argue that the inability to accurately identify such correlation (e.g., between hyperactivity and sugar intake, or between performance on exams and the time of day the exams are held) is one of the most robust shortcomings in human reasoning.[10] People often imagine a correlation between events when no such correlation exists.[11] Jennings, Amadibile, and Ross [1982] argue that illusory correlation can play an important role in the confirmation of false hypotheses, finding that people underestimate correlation when they have no theory of the correlation, but exaggerate correlation and see it where it is not when they have a preconceived theory of it.[12]

Third, confirmatory bias occurs when people selectively collect or scrutinize evidence. One form of "scrutiny-based" confirmatory bias is what we shall call *hypothesis-based filtering*.[13] While it is sensible to interpret ambiguous data according to current hypotheses, people tend to use the consequent "filtered" evidence

10. As Jennings, Amabile, and Ross [1982, p. 212] put it, "even the staunchest defenders of the layperson's capacities as an intuitive scientist . . . have had little that was flattering to say about the layperson's handling of bivariate observation."

11. Chapman and Chapman [1967, 1969, 1971] demonstrate that clinicians and laypeople often perceive entirely illusory correlation among (for instance) pictures and the personality traits of the people who drew the pictures. Stangor [1988] and Hamilton and Rose [1980] also discuss the role of illusory correlation in the context of confirmatory-like phenomena.

12. Similarly, Redelmeier and Tversky [1996] argue illusory correlation may help explain the persistent belief that arthritis pain is related to the weather.

13. Another mechanism can be defined as "positive test strategy": People tend to ask questions (of others, of themselves, or of data) that are likely to be true if their hypothesis is true—without due regard to the fact that they are likely to be true even if the hypothesis is *false*. See Einhorn and Hogarth [1978]; Klayman and Ha [1987]; Beattie and Baron [1988]; Devine, Hirt, and Gehrke [1990]; Hodgins and Zuckerman [1993]; Friedrich [1993]; and Zuckerman, Knee, Hodgins, and Miyake [1995]. We are using this term a bit differently than we suspect psychologists would use it. As far as we know, the term was coined by Klayman and Ha to point out that much of what was put under the rubric of confirmatory bias could indeed be a rational form of hypothesis testing. Fischhoff and Beyth-Marom [1983, pp. 255–256] and Friedrich also point out that if people are fully aware that asking "soft" questions teaches them little about the truth of hypotheses, then no bias has occurred. While we feel research on the positive test strategy needs more careful calibration versus Bayesian updating, we believe that the evidence suggests that people do not fully appreciate how little they have learned about the validity of their hypotheses when asking soft questions. (Mehle, Gettys, Manning, Baca, and Fisher [1981], for instance, show that people with specified hypotheses for observed data tend to overuse such hypotheses to explain the data because they do not have "available" the many unspecified hypothesis that could also explain the data.)

inappropriately as further evidence for these hypotheses. If a student gives an unclear answer to an exam question, it is reasonable for a teacher to be influenced in his evaluation of the answer by his prior perceptions of that student's mastery of the material. However, after assigning differential grades to students according to differential interpretation of comparable answers, it is a mistake to *then* use differential grades on the exam as *further* evidence of the differences in the students' abilities.[14] This sort of error is especially likely when the complexity and ambiguity of evidence requires the use of prior theories when interpreting data and deciding what data to examine.[15]

Finally, one of the main results in our model is confirmation of the conjecture common in the psychological literature that confirmatory bias leads to overconfidence. A vast body of psychological research, separate from research on confirmatory bias, finds that people are prone toward overconfidence in their judgments.[16]

14. Lord, Ross, and Lepper [1979, pp. 2106–2107] note a similar distinction in reflecting on the bias in their experiment discussed above. They note that it is proper for people to differentially assess probative value of different studies according to their current beliefs about the merits of the death penalty. The "sin" is in using their hypothesis-based interpretations of the strength of different studies as further support for their beliefs.

15. We suspect that hypothesis-based filtering is especially important in understanding persistence and strengthening of beliefs in tenuous "scientific" theories. Indeed, Jon Elster drew our attention to an illustration by philosopher of science Karl Popper [1963, pp. 34–35] of confirmatory bias in intellectual pursuits. Popper observed that followers of Marx, Freud, and Adler found "confirmation" everywhere, and described the process by which they strengthened their conviction over time in terms remarkably similar to the process as we've described it based on psychological research:

> Once your eyes were thus opened you saw confirming instances everywhere: the world was full of *verifications* of the theory. Whatever happened always confirmed it . . . The most characteristic element in this situation seemed to me the incessant stream of confirmations . . . As for Adler, I was much impressed by a personal experience. Once, in 1919, I reported to him a case which to me did not seem particularly Adlerian, but which he found no difficulty in analysing in terms of his theory of inferiority feelings, although he had not even seen the child. Slightly shocked, I asked him how he could be so sure. "Because of my thousandfold experience," he replied; whereupon I could not help saying: "And with this new case, I suppose, your experience has become thousand-and-one-fold."
>
> What I had in mind was that his previous observations may not have been much sounder than this new one; that each in its turn had been interpreted in the light of "previous experience," and at the same time counted as additional confirmation.

16. See, e.g., Oskamp [1982], Mahajan [1992], and Paese and Kinnaly [1993]. An early paper that makes this point is Fischhoff, Slovic, and Lichtenstein [1977], who also tested the robustness of overconfidence with monetary stakes rather than reported judgments. No decrease in overconfidence was found relative to the no-money-stakes condition. (As Camerer [1995] notes, there exist *very* few conclusions reached by researchers on judgment that have been overturned when

III. CONFIRMATORY BIAS AND BELIEF FORMATION

Consider two states of the world, $x \in \{A,B\}$, where A and B are two exhaustive and mutually exclusive hypotheses regarding some issue. We consider an agent whose prior belief about x is given by prob $(x = A)$ = prob $(x = B)$ = 0.5, so the agent initially views the two alternative hypotheses as equally likely to be true. In every period $t \in \{1,2,3, \ldots\}$ the agent receives a signal, $s_t \in \{a,b\}$, that is correlated with the true state of the world. Signals received at different times t are independently and identically distributed, with prob $(s_t = a|A)$ = prob $(s_t = b|B)$ = θ, for some $\theta \in (.5,1)$. After receiving each signal, the agent updates his belief about the relative likelihood of $x = A$ and $x = B$.

To model confirmatory bias, we suppose that the agent may misinterpret signals that conflict with his current belief about which hypothesis is more likely. Suppose that, given the signals the agent thinks he has observed in the first $t - 1$ periods, he believes that state A is more likely than state B. Because of his confirmatory bias, the agent may misread a conflicting signal $s_t = b$ in the next period, believing instead that he observes $s_t = a$.

Formally, in every period $t \in \{1,2,3, \ldots\}$ the agent *perceives* a signal $\sigma_t \in \{\alpha,\beta\}$. When the agent perceives a signal $\sigma_t = \alpha$, he believes that he actually received a signal $s_t = a$, and if he perceives $\sigma_t = \beta$, he believes that he actually received a signal $s_t = b$. He updates his beliefs using Bayes' Rule given his (possibly erroneous) perceptions of the signals he is receiving. We assume that with probability $q > 0$ the agent misreads a signal s_t that conflicts with his belief about which hypothesis is more likely, and that the agent always correctly interprets signals that confirm his belief. If he currently believes that Hypothesis A is more likely, then for sure he interprets a signal $s_t = a$ as $\sigma_t = \alpha$, but with probability q he misreads $s_t = b$ as $\sigma_t = \alpha$.

This model of confirmatory bias incorporates several unrealistic simplifying assumptions. For instance, we assume that the severity of the bias summarized by q does not depend on the strength of the agent's beliefs about which of the two states is more likely. It would be reasonable to expect that q is greater if the

---

monetary stakes are added.) There have, however, been criticisms of the evidence in support of overconfidence. See Bjorkman [1994]; Pfeifer [1994]; Tomassini, Solomon, Romney, and Krogstad [1982]; Van Lenthe [1993]; and Winman and Juslin [1993]. We feel, nevertheless, that the evidence makes a strong case for overconfidence. Indeed, see Soll [1996] for evidence that overconfidence *does* extend to ecologically valid domains.

agent's beliefs are more extreme. We conjecture that our qualitative results would continue to hold if we were to relax this assumption. Also, we assume that the agent misreads conflicting evidence as confirming evidence. While we feel that this is often the case, a reasonable alternative model would be to assume instead that the agent merely has a tendency to overlook evidence that conflicts with his beliefs. This model, too, would yield the same qualitative results as our model; intuitively, ignoring the counterhypothesis evidence in a cluster of mixed, but mostly counterhypothesis evidence, is equivalent to misreading the whole cluster as hypothesis-supportive.

The presence of confirmatory bias means that the agent's perceived signals σ_t are neither independently nor identically distributed. Suppose that, after receiving signals $s^{t-1} = (s_1, \ldots, s_{t-1})$ the agent has perceived a sequence of signals $\sigma^{t-1} = (\sigma_1, \ldots, \sigma_{t-1})$ and holds beliefs prob $(x = A | \sigma^{t-1})$. Define

$$\theta^* \equiv \text{prob } (\sigma_t = \alpha | \text{prob } (x = A | \sigma^{t-1}) > 0.5, x = B)$$

$$= \text{prob } (\sigma_t = \beta | \text{prob } (x = B | \sigma^{t-1}) > 0.5, x = A).$$

$$\theta^{**} \equiv \text{prob } (\sigma_t = \alpha | \text{prob } (x = A | \sigma^{t-1}) > 0.5, x = A)$$

$$= \text{prob } (\sigma_t = \beta | \text{prob } (x = B | \sigma^{t-1}) > 0.5, x = B).$$

θ^* and θ^{**} summarize the distribution of the agent's perceived signal σ_t when the agent believes that one hypothesis is more likely than the other; i.e., when prob $(x = A | \sigma^{t-1}) \neq 0.5$. θ^* is the probability that the agent perceives a signal confirming his belief that one hypothesis is more likely when in fact the other hypothesis is true. θ^{**} is the probability that the agent perceives a signal confirming his belief that a hypothesis is more likely when in fact it is true. Because with probability q the agent misreads a signal that conflicts with his beliefs, $\theta^* = (1 - \theta) + q\theta$ and $\theta^{**} = \theta + q(1 - \theta)$. When prob $(x = A | \sigma^{t-1}) = 0.5$, i.e., when the agent believes that the two possible hypotheses are equally likely, the agent does not suffer from confirmatory bias. In this case, he correctly perceives the signal that he receives, and he updates accurately, so $\theta \equiv \text{prob } (\sigma_t = \alpha | \text{prob } (x = A | \sigma^{t-1}) = 0.5, x = A) = \text{prob } (\sigma_t = \beta | \text{prob } (x = B | \sigma^{t-1}) = 0.5, x = B)$.

If $q = 0$, then the agent is an unbiased Bayesian statistician; while if $q = 1$, the agent's first piece of information completely determines his final belief, since he always misreads signals that

conflict with the first signal he receives. More generally, the higher is q, the more extreme is the confirmatory bias.

Suppose that the agent has perceived n_α α signals and n_β β signals, where $n_\alpha > n_\beta$. Because the agent believes he has received n_α a signals and n_β b signals, his updated posterior beliefs are given by

$$\text{prob}\,(x = A | n_\alpha, n_\beta) = \frac{\theta^{n_\alpha - n_\beta}}{\theta^{n_\alpha - n_\beta} + (1 - \theta)^{n_\alpha - n_\beta}}\,.$$

Define

$$\Lambda(n_\alpha, n_\beta) = \frac{\text{prob}\,(x = A | n_\alpha, n_\beta)}{\text{prob}\,(x = B | n_\alpha, n_\beta)}\,.$$

$\Lambda(n_\alpha, n_\beta)$ represents the agent's beliefs in terms of a relative likelihood ratio. Using Bayes' Rule, $\Lambda(n_\alpha, n_\beta) = (\theta^{n_\alpha - n_\beta})/(1 - \theta)^{n_\alpha - n_\beta}$. If $\Lambda(n_\alpha, n_\beta) > 1$, the agent believes that A is more likely than B to be the true state; while if $\Lambda(n_\alpha, n_\beta) < 1$, the agent believes that B is more likely than A. If $\Lambda(n_\alpha, n_\beta) = 1$, the agent believes that the two states are equally likely. The agent's interpretation of an additional signal is biased whenever $\Lambda(n_\alpha, n_\beta) \neq 1$.

In order to identify the effects of confirmatory bias, it is helpful to compare the agent's beliefs with the beliefs of a hypothetical unbiased, Bayesian observer who learns how many α and β signals the agent has perceived, and who knows that the agent suffers from confirmatory bias. Like the agent, the Bayesian observer initially believes that prob $(x = A)$ = prob $(x = B)$ = 0.5, and she has no independent information about whether $x = A$ or $x = B$. This hypothetical observer's beliefs, therefore, reflect the true probability that $x = A$ and $x = B$, given the signals that the agent has perceived.

Define $\Lambda^*(n_\alpha, n_\beta)$ as the Bayesian observer's likelihood ratio of A versus B when she knows that an agent who suffers from confirmation bias has perceived n_α α signals and n_β β signals, where $n_\alpha > n_\beta$. In general, when $q > 0$, the biased agent's likelihood ratio $\Lambda(n_\alpha, n_\beta)$ and the unbiased observer's likelihood ratio $\Lambda^*(n_\alpha, n_\beta)$ are not equal. If $\Lambda(n_\alpha, n_\beta)$ when $n_\alpha > n_\beta$, the agent is *overconfident;* his belief in favor of the hypothesis that $x = A$ is stronger than is justified by the available evidence. Similarly, if $\Lambda(n_\alpha, n_\beta) < \Lambda^*(n_\alpha, n_\beta)$, the agent is *underconfident* in his belief that $x = A$.

In the formal results that we develop below, we assume that, while the unbiased observer knows how many α and β signals the agent has perceived, she does not know the order in which the agent perceived his signals. But when $q > 0$, the order of the agent's perceived signals, if known, would influence a Bayesian observer's beliefs, since the agent's confirmatory bias implies that his perceived signals are not distributed independently. Suppose that the agent has perceived three α signals and two β signals, in which case his beliefs are $\Lambda(n_\alpha = 3, n_\beta = 2) = \theta/(1 - \theta)$. If the Bayesian observer knew the order of the agent's signals, her posterior belief $\Lambda^*(n_\alpha = 3, n_\beta = 2)$ could be less than, greater than, or equal to $\theta/(1 - \theta)$, depending on the order of the signals. Thus, from the perspective of an outside observer, the agent could be overconfident, underconfident, or perfectly calibrated in his beliefs.

Suppose, for example, that the Bayesian observer knew that the agent's sequence of perceived signals was $(\alpha,\alpha,\alpha,\beta,\beta)$. In this case the observer's posterior likelihood ratio is

$$\Lambda^* = \frac{\theta(\theta + q(1 - \theta))^2(1 - \theta)^2}{(1 - \theta)(1 - \theta + q\theta)^2\theta^2}$$

$$= \frac{(\theta + q(1 - \theta))^2(1 - \theta)}{(1 - \theta + q\theta)^2\theta} < \frac{\theta}{1 - \theta}, \ \forall q \in (0,1].$$

Intuitively, the Bayesian observer recognizes the possibility that the agent may have misread his second and third signals, perceiving that they supported the hypothesis that $x = A$ when in fact one or both may have supported the hypothesis that $x = B$. Therefore, the Bayesian observer is less convinced that $x = A$ than the agent, who is overconfident in his belief. More generally, an observer who knows that a biased agent has always believed in his current hypothesis should judge the agent to be overconfident in his belief, since there is a positive probability that the agent has misread signals that are counter to his favored hypothesis. An observer who knows that a teacher has always believed that Bart is smarter than Marta should recognize that the teacher's confirmatory bias may have led him to misread evidence that Marta is in fact smarter.

Alternatively, suppose that the Bayesian observer knew that the agent's sequence of perceived signals was $(\beta,\beta,\alpha,\alpha,\alpha)$. Now the

observer's posterior likelihood ratio is

$$\Lambda^* = \frac{(1 - \theta)(1 - \theta + q\theta)\theta^3}{\theta(\theta + q(1 - \theta))(1 - \theta)^3}$$

$$= \frac{(1 - \theta + q\theta)\theta^2}{(\theta + q(1 - \theta))(1 - \theta)^2} > \frac{\theta}{1 - \theta} \, , \ \forall q \in (0,1].$$

In this case, the Bayesian observer believes that the agent may have misread his second signal, perceiving that it supported the hypothesis that $x = B$ when in fact it may have supported the hypothesis that $x = A$. Thus, the Bayesian observer believes that there is a greater likelihood that $x = A$ than the agent, who is *under*confident in his belief. More generally, an observer who knows that a biased agent only recently came to believe in his current hypothesis after long believing in the opposite hypothesis should judge the agent to be underconfident in his belief, since the agent may have misread one or more signals that support his current hypothesis when he believed the opposite. An observer who knows that a teacher initially thought that Bart was smarter than Marta, but eventually started to believe that it was slightly more likely that Marta was smarter than Bart, should conjecture that the teacher is underconfident about his new hypothesis. When the teacher believed that Bart was smarter than Marta, he may have misinterpreted signals that Marta was smarter. The fact that the teacher came to believe that Marta was smarter *despite* his initial bias toward believing that Bart was smarter indicates that the evidence is very strong that Marta is smarter.[17]

The preceding examples illustrate how information about the order of the agent's signals would significantly influence an

---

17. As a discussant for this paper, Roger Lagunoff made an interesting suggestion that is especially relevant for the examples we are discussing here. In our model, once the agent interprets a signal, he never goes back and reinterprets it—even if he later changes his hypothesis about the world. Hence we are not capturing a form of belief updating we sometimes observe: when somebody (finally) comes around to change his world-view that he held for quite a while, he sometimes experiences an epiphany whereby he goes back and reinterprets previous evidence in light of his new hypothesis, realizing that "the signs were there all along." This suggests a model in which an agent is biased in interpreting not just the next signal, but all past signals, as supporting his current hypothesis about the world. While we suspect there is some truth to this, we don't believe that people fully retroactively rebias themselves in this way. (We have found no psychological evidence about this one way or another.) While such an alternative model would rule out the possibility of "underconfidence" for recent converts, it would leave all the predictions regarding overconfidence discussed in the remainder of the paper qualitatively the same, and magnify the magnitude of our results (and simplify the proofs).

outside observer's judgment about whether, and in what direction, the agent's beliefs were biased. Nevertheless, for the remainder of the paper we assume that an outside observer only knows the number of α and β signals that the agent has received, and not the order in which he received them. This assumption enables us to identify whether, on *average,* the agent is over- or underconfident. This appears to be the question that the psychological literature addresses; presumably, it is also of interest to economists.

Clearly, if $q = 0$, then $\Lambda^*(n_\alpha,n_\beta) = \Lambda(n_\alpha,n_\beta)$. When $q > 0$, however, Proposition 1 establishes that $\Lambda^*(n_\alpha,n_\beta) < \Lambda(n_\alpha,n_\beta)$. That is, when the agent perceives that a majority of his signals support (say) Hypothesis A, he believes in A with higher probability than is warranted.[18]

PROPOSITION 1. Suppose that $n_\alpha > n_\beta$ and $n_\alpha + n_\beta > 1$. Then $\Lambda^*(n_\alpha,n_\beta) < \Lambda(n_\alpha,n_\beta)$.

Proposition 1 establishes that an agent who suffers from confirmatory bias will be overconfident in his belief about which state is most likely.

An observer who knows the agent's beliefs cannot usually observe the exact sequence of the agent's perceived signals. Therefore, the observer's judgment about whether the agent is under- or overconfident depends on her belief regarding the likelihood of the different possible sequences of signals. Proposition 1 establishes that overconfidence is the dominant force. The intuition for this result is fairly straightforward: if you cannot directly observe the agent's past beliefs, but you know that he now believes in Hypothesis A, you should surmise that, on average, he spent more time in the past believing Hypothesis A than Hypothesis B. Consequently, you should surmise that, on average, the agent misread more signals while believing in Hypothesis A—contributing to overconfidence—than he misread while believing in Hypothesis B—contributing to underconfidence. Proposition 1 hinges to some extent on our assumption that the agent receives signals that are the same strength in every period. We believe that (far more complicated) versions of Proposition 1 hold in more general models, but we show in Appendix 1 that *under*confidence is sometimes possible when the agent's signals are of different strengths in different periods.

18. All proofs are in Appendix 2. Because our model is entirely symmetric, we shall for convenience present all results and much of our discussion solely for the case where A is perceived as more likely.

Proposition 1 shows that when the agent believes that the state is $x = A$ with probability $\mu > 0.5$, the true probability that the state is $x = A$ is less than μ. Interestingly, the true probability that A is the true state *may be less than* 0.5, meaning that B is more likely than A. The possibility that the agent may suffer not merely from overconfidence, but also from "wrongness," arises when the agent's confirmatory bias is severe and he has perceived at least two signals in favor of each hypothesis.

To see the intuition for this result, suppose that the agent has since his first signal $s_1 = a$ believed that Hypothesis A is more likely than B, but that he nevertheless has perceived two signals $\sigma_t = \sigma_t = \beta$ at two times $t, t' > 1$. If the agent's confirmatory bias is severe (i.e., $q \approx 1$), only his first perceived signal in favor of A provides true evidence that $x = A$. Once the agent believes that A is true, his confirmatory bias predisposes him to perceive that subsequent signals support this belief, and, therefore, additional signals in favor of A are not very informative. But, because the agent's two perceived signals in favor of B conflict with what he believes—that $x = A$ is more likely—they reflect *actual* signals in favor of B. Thus, although the agent has always believed that $x = A$ is more likely, he has effectively received only one signal in favor of A and two signals in favor of B. In this case the agent's belief that $x = A$ represents extreme overconfidence; if he had correctly interpreted evidence, he would believe that $x = B$ is more likely.

It is, of course, possible that hypothesis A is *more* likely than the agent realizes if he first perceives a signal $s_1 = b$, falsely reads a's as b's for a while, and only later perceives enough α's to come to believe in A. And it is true that getting more true a's than true b's implies that Hypothesis A is more likely. Yet, it can be shown that these possibilities may be far less likely than the cases leading to extreme overconfidence, so that the net effect that is more likely that B is true than that A is true if the agent believes in A with mixed evidence.

For example, suppose that the agent has perceived seven signals, four α's and three β's. Given these signals, the agent's posterior beliefs are $\Lambda(n_\alpha = 4, n_\beta = 3) = \theta/(1 - \theta) > 1$; the agent believes that the state $x = A$ is more likely. Meanwhile, the true likelihood ratio is $\Lambda^*(n_\alpha = 4, n_\beta = 3) =$

$$\frac{\begin{array}{c}(1 - \theta)^3[8\theta^4 + 8\theta^3\theta^{**} + 7\theta^2\theta^{**2} + 5\theta\theta^{**3}] \\ + (1 - \theta)^2[\theta^3\theta^{**}\theta^* + 4\theta^4\theta^*] + 2\theta^4(1 - \theta)\theta^{*2}\end{array}}{\begin{array}{c}\theta^3[8(1 - \theta)^4 + 8(1 - \theta)^3\theta^* + 7(1 - \theta)^2\theta^{*2} + 5(1 - \theta)\theta^{*3}] \\ + \theta^2[(1 - \theta)^3\theta^{**}\theta^* + 4(1 - \theta)^4\theta^{**}] + 2\theta(1 - \theta)^4\theta^{**2}\end{array}} .$$

Suppose that $\theta = .75$. Then the agent's posterior likelihood ratio is $\Lambda(4,3) = 3$. Suppose further that $q = .95$, and therefore the agent suffers from severe confirmatory bias. Then, the true likelihood ratio is $\Lambda^*(4,3) = .63$, and therefore $x = B$ is more likely to be the true state, despite the agent having perceived more α signals than β signals.

Indeed, it turns out to be the case that when confirmatory bias is very severe *and* the signals are very informative, then whenever you observe the agent believing in Hypothesis A and having perceived two or more β signals, then you should assume that it is more likely that B is true than A. We formalize this in Proposition 2. Let $\Lambda^*(n_\alpha, n_\beta | q, \theta)$ be the appropriate beliefs as a function of q and θ. Then

PROPOSITION 2. For $n_\alpha > n_\beta$ and $n_\beta \leq 1$, $\lim_{\epsilon \to 0} \Lambda^*(n_\alpha, n_\beta | 1 - \epsilon, 1 - \epsilon) > 1$. For all $n_\alpha > n_\beta \geq 2$, $\lim_{\epsilon \to 0} \Lambda^*(n_\alpha, n_\beta | 1 - \epsilon, 1 - \epsilon) < 1$.

That is, for θ and q both very close to 1, when the agent has perceived one or fewer β signals and believes in Hypothesis A, she is probably correct (though overconfident) in her beliefs; when the agent has perceived two or more β signals and believes in Hypothesis A, she is probably *in*correct in her beliefs—Hypothesis B is more likely to be true.

We emphasize that the very premise of the proposition means that the situations to which it applies are uncommon; when both q and θ are close to 1, the probability of perceiving anything besides a sequence of signals favoring the correct hypothesis is small. Therefore, Proposition 2 tells us about a very low-probability event. In our example with seven signals, $q = .95$, and $\theta = .75$, the probability that the signals are sufficiently mixed that the agent is probably wrong is a little more than one-half percent.

While we do not know more generally the highest probability with which the agent can be wrong, some calibrations illustrate that it *can* be relatively likely that the agent ends up with beliefs that a Bayesian observer would deem probably wrong. Tables I–IV display, for various values of n, θ, and q, the probability that $\Lambda^* < \eta$ and $\Lambda > 1$ or $\Lambda^* > 1/\eta$ and $\Lambda < 1$, where η represents different thresholds for how wrong the agent is. Table entries are in percentage terms (rounded to the nearest percent), with rows corresponding to different values of q and columns to different values of θ. (Dashes indicate an entry *exactly* equal to zero.)[19]

For instance, with $\theta = .6$ and $q = .5$, the probability that the

19. The entries in Tables I–IV reflect direct calculations (performed by computer) of the probabilities in question.

TABLE I
PROBABILITY OF "WRONGNESS," $n = 50$, $\eta = 1$

	θ			
q	0.6	0.7	0.7	0.9
.1	—	—	—	—
.2	12	2	0	—
.3	21	9	1	0
.4	29	15	5	1
.5	27	18	10	3
.6	33	22	12	5
.7	27	21	15	7
.8	33	24	15	8
.9	21	17	12	9

TABLE II
PROBABILITY OF "WRONGNESS," $n = 50$, $\eta = \frac{1}{2}$

	θ			
q	0.6	0.7	0.7	0.9
.1	—	—	—	—
.2	—	—	—	—
.3	10	5	1	—
.4	15	13	5	1
.5	19	18	9	3
.6	16	18	12	5
.7	18	21	13	7
.8	11	16	15	7
.9	4	8	12	6

TABLE III
PROBABILITY OF "WRONGNESS," $n = 50$, $\eta = \frac{1}{9}$

	θ			
q	0.6	0.7	0.7	0.9
.1	—	—	—	—
.2	—	—	—	—
.3	—	—	—	—
.4	—	5	—	—
.5	3	12	7	1
.6	3	14	11	4
.7	2	11	12	6
.8	1	6	12	7
.9	0	4	7	6

TABLE IV
PROBABILITY OF "WRONGNESS," $n = 7$, $\eta = 1$

	θ			
q	0.6	0.7	0.7	0.9
.1	—	—	—	—
.2	—	—	—	—
.3	—	—	—	—
.4	—	—	—	—
.5	—	—	—	2
.6	—	5	3	1
.7	3	3	2	5
.8	10	8	6	3
.9	3	2	2	1

agent has beliefs after 50 signals that the observer would deem probably wrong is about 27 percent. The probability in this same case that his beliefs will lead the observer to believe in the other hypothesis with at least probability ⅔ is 19 percent, and the probability that the observer would believe in the hypothesis opposite to the agent's with at least 9/10 probability is about 3 percent.[20]

In the example above and in Proposition 2, the agent can be wrong in her beliefs. Even more surprising, perhaps, the true probability that A is the correct hypothesis need not be monotonically increasing in the proportion of α signals the agent perceives. Continue to assume that the agent has received seven signals, but now suppose that five support $x = A$ and two support $x = B$. Then, because $\theta = .75$, the agent's posterior likelihood ratio is $\Lambda(5,2) = 27 > \Lambda(4,3)$. Meanwhile, the true likelihood ratio is

$$\Lambda^*(n_\alpha = 5, n_\beta = 2)$$

$$= \frac{(1 - \theta)^2[7\theta^2\theta^{**3} + 9\theta\theta^{**4} + 4\theta^3\theta^{**2}] + (1 - \theta)\theta^3\theta^{**2}\theta^*}{\theta^2[7(1 - \theta)^2\theta^{*3} + 9(1 - \theta)\theta^{*4} + 4(1 - \theta)^3\theta^{*2}] + \theta(1 - \theta)^3\theta^{*2}\theta^{**}} .$$

20. Readers may note that these probabilities generally increase in q and then decrease, with probability about 0 for $q = 0$ and $q = 1$. But they are not single-peaked in q. This is because there are two factors at work in determining the influence of q on the probability. As q increases, the probability that the agent will end up with close-to-even mixes of α and β signals decreases continuously. But because an increase in q increases the likelihood that any given combination of α's and β's involves the agent being probabilistically wrong, there will be at certain points discrete jumps upward in the likelihood of wrongness for some values of q. The result is an extremely poorly behaved function.

Maintaining the assumption that $q = .95$, $\Lambda^*(5,2) = .62 <$ $\Lambda^*(4,3) = .63$. Therefore, the relative likelihood that the true state is $x = A$ versus $x = B$ is smaller if the agent perceives that five out of seven signals support $x = A$ than if he perceives that only four out of seven signals support $x = A$.

While seemingly counterintuitive, this result reflects the fact that the agent is more likely to have perceived (truly informative) signals $\sigma_j = \beta$ that conflict with a belief that $x = A$ when he has perceived only two signals in favor of B than when he has perceived three signals in favor of B. Intuitively, the agent is more likely to have believed for many periods that $x = A$ in the former case than in the latter case. Put differently, the agent is *less likely* to have perceived (truly informative) signals $\sigma_j = \alpha$ that conflict with a belief that $x = B$ when he has perceived only two signals in favor of B than when he has perceived three signals in favor of B.

The preceding examples illustrate that an agent who suffers from confirmatory bias may believe that one of the two possible states is more likely than the other when in fact the reverse is true. Nevertheless, Proposition 3 shows that a Bayesian observer who knows only that a biased agent believes that $x = A$ is more likely than $x = B$ will herself believe that $x = A$ is more likely. Therefore, an agent who suffers from confirmatory bias will "on average" correctly judge which of the two possible states is more likely, though, as Proposition 1 establishes, he will always be overconfident in his belief.

Define $\Lambda^*(n)$ as the likelihood ratio of a Bayesian observer who knows that a confirmatory agent has perceived a total of n signals, and knows that $n_\alpha > n_\beta$, but does not know the exact values of n_α and n_β. That is, the observer knows only that the agent believes A is more likely than B, but observes nothing about the strength of his beliefs. Then

PROPOSITION 3. For all n, $\Lambda^*(n) > 1$.

In light of the above examples where the agent may be wrong, the simple generality of Proposition 3 may seem surprising. It is reconciled with the examples by observing that the agent suffers from "wrongness" only when his confirmatory bias is very severe, meaning that q is close to 1, and yet he has perceived mixed signals about which state is more likely. But the agent is unlikely to receive mixed signals when his confirmatory bias is strong, because each signal σ_t will tend to mirror σ_1.

IV. Beliefs after an Infinite Number of Signals

A fully Bayesian agent—for whom $q = 0$—will after an infinite number of signals come to believe with near certainty in the correct hypothesis. We now investigate the implications of confirmatory bias in the limit as an agent receives an infinite number of signals.

We begin with definitions and a lemma that will help to analyze this question. Suppose that the agent has thus far received $m = n_\alpha - n_\beta > 0$ more perceived signals in support of Hypothesis A than in support of Hypothesis B. Suppose further that, as long as $n_\alpha > n_\beta$, prob $(\sigma_t = \alpha) = \gamma$. Note that $\gamma = \theta^*$ if B is true, and $\gamma = \theta^{**}$ if A is true. We wish to consider some preliminary results that hold in either case. We define $p(m,\gamma)$ as the probability that there exists some time in the future when the agent will have received an equal number of α and β signals. (At that time the agent's posterior belief is the same as his prior belief, prob $(x = A) = 0.5$.) We have the following lemma, which is a restatement of a well-known result from Feller [1968, pp. 344–347).

LEMMA 1. For all $m > 0$, $\gamma \geq 0.5$, $p(m,\gamma) = [(1 - \gamma)/\gamma]^m$. For $\gamma \leq 0.5$, $p(m,\gamma) = 1$.

We define P_W as the probability that the agent, beginning with the prior belief prob $(x = A) = 0.5$, comes to believe with certainty in the wrong hypothesis after receiving an infinite number of signals.[21] That is, P_W is the probability that, although the true state is $x = A$, the agent instead comes to believe irreversibly, with near certainty, that $x = B$. Proposition 4 characterizes P_W as a function of q and θ.

PROPOSITION 4. If $q > 1 - 1/(2\theta)$, then

$$P_W = \frac{(1 - \theta) \cdot (1 - (1 - \theta^*)/\theta^*)}{(1 - (1 - \theta) \cdot ((1 - \theta^*)/\theta^*) - \theta((1 - \theta^{**})/\theta^{**}))} > 0.$$

If $q \leq 1 - 1/(2\theta)$, then $P_W = 0$.

When $q > 1 - 1/(2\theta)$, $\theta^* = (1 - \theta) + q\theta > 0.5$. When $\theta^* > 0.5$, then once the agent comes to believe that the wrong hypothesis about x is more likely, he is consequently more likely to receive a

---

21. Formally, $P_W \equiv$ prob $(\forall k > 0$ and $\forall \epsilon > 0$, $\exists n^*$ such that prob $(n_\alpha - n_\beta > k$ for all $n > n^*) > 1 - \epsilon)$.

TABLE V
PROBABILITY OF BELIEVING IN WRONG HYPOTHESIS AFTER OBSERVING AN INFINITE
NUMBER OF SIGNALS

q	$\theta = .6$	$\theta = .667$	$\theta = .75$	$\theta = .9$
.25	18	—	—	—
.333	26	12	—	—
.5	34	24	13	2
.75	38	31	22	8

signal σ_t that *confirms* this incorrect belief than he is to receive a signal that *conflicts* with this incorrect belief. This guarantees that there is a positive probability that the agent will never overturn his incorrect hypothesis, and in fact come to believe more and more strongly in that wrong hypothesis. Conversely, if $q < 1 - 1/(2\theta)$, then $\theta^* < .5$, which guarantees that the agent will, every time he comes to believe the wrong hypothesis is more likely, eventually come to abandon that belief. This in turn implies that the agent will repeatedly come to believe the correct hypothesis is more likely; and since $\theta^{**} = \theta + q(1 - \theta) > \theta > .5$, he will eventually come to believe in it with near certainty.

The proposition shows that, despite receiving an infinite number of signals, the agent may become certain that the incorrect hypothesis is in fact true.[22] This occurs when the agent's confirmatory bias is sufficiently severe. To illustrate the magnitude of P_W, Table V displays P_W for various values of θ and q. Table entries are in percentage terms (rounded to the nearest percent), with rows corresponding to different values of q and columns to different values of θ. (Dashes indicate an entry *exactly* equal to zero.)

For example, suppose that $q = 0.5$ and $\theta = .75$. Then $P_W = \frac{7}{52}$, meaning that approximately 13 percent of the time the agent will eventually come to believe with certainty in the wrong hypothesis. As the quality of the agent's true signal worsens he is more likely to believe with certainty in the wrong hypothesis. Indeed, a corollary to Proposition 4 is that, fixing any $q > 0$, $\lim_{\theta \to 1/2} P_W = \frac{1}{2}$.

We now investigate the related question of when the agent will maintain an incorrect initial belief. To do so, we relax our

22. It is straightforward to show that the agent becomes certain that the *correct* hypothesis about the state of the world is true with complementary probability. Therefore, after an infinite number of signals the agent will believe that *one* of the hypotheses is certainly true.

assumption that the agent initially believes that each state x is equally likely and suppose instead that the agent initially believes that the wrong hypothesis is more likely to be true. For example, if $x = B$ is the true state of the world, then the agent initially believes prob $(x = A) = \mu > 0.5$. Crucially, we assume that this belief arose from signals that are independent of the new signals that the agent receives, which are distributed as outlined above.

Given the assumption that the signals are independently distributed and ignoring integer problems, these prior beliefs can be interpreted as if the agent has already received D more signals supporting the incorrect hypothesis, where

$$\mu = \theta^D/[\theta^D + (1 - \theta)^D].$$

This formula implicitly defines a function $D(\mu)$. The agent must receive $D(\mu)$ more conflicting signals σ_t than confirming signals in order to reach a posterior belief that the two possible states of the world, A and B, are equally likely.

We define $P_W(\mu)$ as the probability that the agent, beginning with the prior belief $\mu > 0.5$ that the wrong hypothesis about the state of the world is true, comes to believe with certainty in the wrong hypothesis after receiving an infinite number of signals.[23]

PROPOSITION 5. Choose any $\epsilon > 0$ and any $\mu > 0.5$. Then
  (i) For all $\theta \in (0.5,1)$, there exists $q > 0$ such that $PW(\mu) > 1 - \epsilon$.
  (ii) For all $q > 0$, there exists $\theta > 0.5$ such that $PW(\mu) > 1 - \epsilon$.

Proposition 5 says that an agent who begins with an arbitrarily small bias in the direction of the incorrect hypothesis will almost surely maintain his belief in this hypothesis when either of two conditions is satisfied. First, and not very surprisingly, this will occur when the agent is subject to severe confirmatory bias. When q is very close to 1, then the agent almost never receives signals that conflict with his initial belief, and therefore it is not surprising that this belief is rarely overturned. Second, and somewhat more surprisingly, the agent almost surely maintains his incorrect belief provided that his true signals are very weak, meaning that θ is very close to 0.5. This result does not depend on the level of confirmatory bias, so long as $q > 0$. This result means that if the agent receives only very weak feedback from his environment and is subject to any confirmatory bias, he almost

23. $P_W(0.5) = P_W$.

never overcomes any initial beliefs that are significantly incorrect, and in fact comes to believe that the incorrect hypothesis is certainly true. While one should not overinterpret the second result in Proposition 5—we can question whether agents really pay attention to such weak feedback—the conclusion is nevertheless very striking. Propositions 4 and 5 show that an infinite sequence of signals will not necessarily lead people to overcome erroneous beliefs; rather, people may simply become more and more confident in those erroneous beliefs.

Table VI displays $P_W(\mu)$ for various values of θ, q, and μ. If θ and μ are chosen in such a way that $D(\mu)$ is an integer, and $q > 1 - 1/2\theta$, it follows from Lemma 1 that

$$P_W(\mu) = \left(1 - \left[\frac{1 - \theta^*}{\theta^*}\right]^{D(\mu)}\right) + \left(P_W(0.5)\left[\frac{1 - \theta^*}{\theta^*}\right]^{D(\mu)}\right) > 0.$$

Table entries are in percentage terms (rounded to the nearest percent), with rows corresponding to different values of q and columns to different values of θ and the prior belief μ. (Dashes indicate an entry *exactly* equal to zero.)

For example, suppose that $\theta = .551$ and $\mu = .6$. In this case $D(\mu) = 2$, meaning that the agent must receive two more signals that conflict with rather than confirm his prior belief in order to believe that the states A and B are equally likely. But if $q = .333$, there is nearly an 80 percent chance that the agent will never overturn his incorrect prior belief. Clearly, learning does not necessarily lead the agent to correctly identify the true state.

V. Confirmatory Bias in a Principal-Agent Model

Confirmatory bias is likely to influence economic behavior in many different arenas. In this section we develop a simple

TABLE VI
Probability of Maintaining an Incorrect Prior Belief after Receiving an
Infinite Number of Signals

q	$\theta = .6, \mu = .6,$ $D(\mu) = 1$	$\theta = .551, \mu = .6,$ $D(\mu) = 2$	$\theta = .75, \mu = .75,$ $D(\mu) = 1$	$\theta = .634, \mu = .75,$ $D(\mu) = 2$
.25	32	67	—	24
.33	51	79	—	56
.5	72	92	48	85
.75	89	99	82	98

illustrative model of a principal-agent relationship, a context where we think confirmatory bias is likely to be important. The premise of the model is that an agent may take inappropriate actions not solely because of intentional misbehavior—moral hazard—but also because of unintentional errors arising from confirmatory bias. Specifically, because an agent who suffers from confirmatory bias will be overconfident in his judgment about how likely various actions are to pay off, he may be prone to taking actions that are riskier and more "extreme" than is optimal for the principal. Such overconfidence seems to reflect the intuition among some researchers: at a conference one of the authors attended, a leading economist conjectured that bad investment decisions by businesses in Eastern Europe receiving bank loans were more often the result of overconfidence by borrowers than of intentions to mislead banks. Even more directly along the lines of our model, Wood [1989] asserts that money managers become more confident in their investment decisions as they gather more information—even when the quality of their investment decisions is not improved.

A principal who is aware of an agent's confirmatory bias will wish to design incentives that both cause the agent to internalize the negative consequences of bad choices *and* prevent decisions based on good-faith overconfidence. In particular, incentives that lead the agent to collect a lot of information may not be optimal if the agent suffers from severe confirmatory bias and, hence, becomes more overconfident as he collects more information. The principal may therefore wish to mute the agent's incentives relative to what would be optimal in the absence of confirmatory bias.

While an exhaustive analysis of the effect of confirmatory bias on agency relationships is beyond the scope of this paper, we now develop a simple illustrative model along these lines. Suppose that a principal hires an agent to allocate initial wealth $W = 1$ between the different investments in the set $I = \{I_A, I_B, I_C\}$. The investment I_C is risk-free; it always yields a gross return $r(I_C) = 1$. Investments I_A and I_B, on the other hand, are risky; their returns depend on the state of nature $x \in \{A,B\}$. Conditional on the state x, the gross returns from I_A and I_B are $r(I_A|A) = r(I_B|B) = R \in (1,2)$ and $r(I_A|B) = r(I_B|A) = 0$. Let $u(\cdot)$ be the principal's Von Neumann-Morgenstern utility function for money. Because the principal may be risk-averse, we assume that $u' > 0$ and $u'' \leq 0$. Consistent with the model that we developed in the previous sections, the

principal and the agent cannot observe the state x, and they hold a common prior belief that prob $(x = A)$ = prob $(x = B)$ = 0.5. Hence, if the agent learns nothing more about the true state, the optimal investment is in the riskless investment I_C; if he learns sufficiently more—generating beliefs sufficiently different from .5—he will perceive it as optimal to invest some money in one of the two risky investments.[24]

Before choosing how to invest the principal's wealth, the agent has the opportunity to observe informative signals about the true state, although the agent's confirmatory bias may lead him to misinterpret these signals. We assume that the signals that the agent receives, and the way he perceives these signals, accord with the model in the previous sections.

For both analytic ease and to highlight the role of confirmatory bias, we abstract away from the usual moral-hazard concerns: we assume that the agent costlessly observes signals about the state x and expends no effort when making decisions on the principal's behalf. Under this assumption, an arbitrarily small incentive to identify the true state would lead the agent to observe an infinite number of signals, after which he would believe that he could identify the true state with near certainty. Furthermore, to abstract away from issues of optimal risk-sharing between the agent and the principal, we assume that the agent is (nearly) infinitely risk-averse. Therefore, the principal must offer the agent a nearly constant wage.[25] Under these assumptions, the

24. While the language and notation suggest that we are referring to well-defined investment portfolios (e.g., three different bonds), we mean for the model to apply as well to internal organizational incentives to pursue ambiguously defined projects. Indeed, this alternative interpretation may better fit the formal model in some respects. Note that it is crucial to our analysis that the agent cannot or does not merely report his beliefs to the principal, but rather implements a strategy himself based on his beliefs. If the principal knew the agent's beliefs and the extent of his confirmatory bias, she could form her own beliefs about the true state of the world and then directly choose the action that would maximize her expected payoff.

25. These assumptions raise a subtle point. If the agent anticipated gathering an infinite number of signals, he would be willing to accept a contract that yielded a payoff that depended on the outcome of a risky investment, even if he were infinitely risk-averse. This is because the agent would anticipate being able to identify the true state with virtual certainty. But, if the hypothesis of the first part of Proposition 4 is satisfied, the agent will be overconfident in his judgment after observing an infinite number of signals. Therefore, such a contract would impose more risk on the agent—and yield a lower expected utility—than he anticipated. We assume here that the principal cannot exploit the agent's confirmatory bias by convincing him to sign a contract that yields an expected payoff that is less than the agent's reservation payoff. This assumption means that the principal cannot use the agent as a "money pump," and its empirical validity deserves investigation. The issue of whether to focus analysis on "efficiency contracts" rather than "money-pumping contracts" is a more general one that is likely to arise in

incentives that the principal offers to the agent affect the principal's payoff only through their effect on the investment decision that the agent makes. The principal does not need to compensate the agent for gathering information, and she cannot transfer risk to the agent.

These assumptions permit us to focus on two polar cases. In the first case, the principal gives the agent no incentive to collect information, and the agent allocates the principal's wealth with no information beyond his prior belief. In the second case, the principal gives the agent an arbitrarily small incentive to collect information, and the agent allocates the principal's wealth after observing an infinite number of signals and coming to believe with virtual certainty that he has identified the true state.

Suppose first that the principal offers the agent no incentive to gather information. Because $R < 2$ and the principal's prior belief is prob $(x = A) = 0.5$, it is optimal for the principal to direct the agent to invest all of the principal's wealth in the risk-free asset I_C. Now suppose that the principal offers the agent a small incentive to identify the true state. For instance, suppose that the principal offers to pay the agent an arbitrarily small fraction of the principal's gross return. Because it is free for the agent to collect information, he would observe an infinite number of signals and come to believe that he could identify the true state with certainty. These extreme beliefs would lead the agent to allocate all of the principal's wealth to one of the risky assets. If, for example, the agent thought that A was surely the true state, he would allocate all of the principal's wealth to asset I_A.[26]

If $q \leq 1 - 1/(2\theta)$ (i.e., if confirmatory bias is sufficiently weak), Proposition 4 establishes that the agent will eventually identify the true state with near certainty if he collects enough signals. Hence, under our assumption that it is costless for the agent to gather information and that in every state of the world one or the other of the "risky" investments is optimal, it would then be optimal for the principal to offer a contract that would lead the agent to collect an infinite number of signals, and the principal would receive a payoff $u(R) > u(1)$.

If $q > 1 - 1/(2\theta)$, on the other hand, Proposition 4 establishes

---

developing formal models of incentives for boundedly rational agents. See, for instance, O'Donoghue and Rabin [1997], who study incentive design for agents who irrationally procrastinate, and who discuss various rationales for focusing on efficiency contracts.

26. We assume that short-selling is impossible, so the agent could not allocate more than \$1 to asset α by, for instance, selling investment I_C short.

that the agent's (completely confident) belief about the true state is wrong with positive probability. It can be shown that he *correctly* identifies the true state with probability

$$\mu^*(\theta,q) = \frac{\theta[2(\theta + q(1 - \theta)) - 1](1 - \theta + q\theta)}{q[1 - 2(1 - q)\theta(1 - \theta)]}.$$

Since the agent is fully confident that he has identified the true state, he invests all of the principal's wealth in the risky asset he believes is most profitable, so the principal's payoff is $\mu^*(\theta,q)u(R) + (1 - \mu^*(\theta,q))u(0)$. Note that $\mu^*(\theta,q)$ is increasing in θ and decreasing in q, meaning that the agent identifies the true state with higher probability when he receives more informative signals and lower probability when his confirmatory bias is more severe.

The principal does not want the agent to become "informed" when $u(1) \geq \mu^*(\theta,q)u(R) + (1 - \mu^*(\theta,q))u(0)$. Define $\underline{\mu}$ as satisfying $u(1) = \underline{\mu}u(R) + (1 - \underline{\mu})u(0)$; if the agent correctly identifies the true state with probability $\underline{\mu}$ after observing an infinite number of signals, the principal is just willing for the agent to become informed about the state x. Because $\mu^*(\theta,q) \geq \theta$, the principal always offers the agent an incentive to become informed if $\theta \geq \underline{\mu}$. That is, if the principal prefers all her money invested in a risky investment based on just one signal to having all her money invested in the risk-free investment, she will provide incentives to the agent. When $\theta < \underline{\mu}$, on the other hand, we have Proposition 6.

PROPOSITION 6. Suppose that $\theta < \underline{\mu}$.

  (i) There exists $q^* \in (1 - 1/2\theta,1]$ such that the principal does not offer the agent an incentive to become informed about the state x if and only if $q \geq q^*$.

  (ii) For any $q \in [0,1]$, there exists $\theta^* \in (0.5,\underline{\mu}]$ such that the principal does not offer the agent an incentive to become informed about the state x if and only if $\theta \leq \theta^*$.

When $\theta < \underline{\mu}$, the principal does not want the agent to observe signals about the state x either if confirmatory bias is very severe or if the agent receives very weak signals. In either case, there is a strong possibility that an "informed" agent would erroneously identify the true state, although the agent himself would overconfidently believe that he could identify the true state with near certainty. If the agent's overconfidence is sufficiently severe, the principal prefers not to offer the agent any incentive to become informed, in which case the agent will invest the principal's wealth in the riskless asset.

The principal's degree of risk aversion also influences whether or not he wants the agent to observe signals about the state x. While Proposition 6 shows that there are conditions where even a risk-neutral principal eschews incentives for the agent, the principal is more bothered by the overconfidence when she is more risk-averse, in the usual sense defined by Pratt [1964]. Indeed, whenever confirmatory bias is severe enough that the agent might be wrong even after gathering an infinite number of signals, a principal who is sufficiently risk-averse will prefer not to offer her agent any incentive to become informed. We formalize this idea in Proposition 7.

PROPOSITION 7. Suppose that $\theta \in (0.5,1)$, $q \in (1 - 1/2\theta,1]$, and $u(\cdot)$ is a Von Neumann-Morgenstern utility function $u(\cdot)$ satisfying $u' > 0$, $u'' \leq 0$.

(i) Suppose that $u(1) \leq \mu^*(\theta,q)u(R) + (1 - \mu^*(\theta,q))u(0)$. Then there exists a function $g(\cdot)$ such that $g' > 0$, $g'' \leq 0$, and $g(u(1)) \geq \mu^*(\theta,q)g(u(R)) + (1 - \mu^*(\theta,q))g(u(0))$.

(ii) Suppose that $u(1) \geq \mu^*(\theta,q)u(R) + (1 - \mu^*(\theta,q))u(0)$. Then for any function $g(\cdot)$ such that $g' > 0$, $g'' \leq 0$, $g(u(1)) \geq \mu^*(\theta,q)g(u(R)) + (1 - \mu^*(\theta,q))g(u(0))$.

(iii) In both (i) and (ii), $v(\cdot) = g(u(\cdot))$ is a Von Neumann-Morgenstern utility function that represents preferences that are globally more risk-averse than those represented by $u(\cdot)$.

Suppose that Marta, whose preferences are represented by $u(\cdot)$, wishes to give her agent the incentive to become informed about x. The proposition establishes that there exist preferences that are globally more risk-averse than Marta's under which a principal would prefer not to give her agent the incentive to become informed. Furthermore, if Marta does not wish to give her agent an incentive to become informed about the state x, then any principal who is globally more risk-averse than Marta would also choose not to offer incentives to an identically biased agent facing the same investment decision.

The preceding analysis reflects an assumption that a biased agent who feels he is fully informed will invest all of the principal's wealth in a single risky asset. The agent will pursue such a strategy if, for example, the principal offers the agent a fixed share of the principal's gross investment return. Our analysis assumes, of course, that the principal cannot directly contract on decisions,

only on returns. But it also implicitly assumes that the principal cannot punish the agent for having too high an expected return. If she could, then she might wish for the agent to gather some information—and then provide incentives such that the (overconfident) agent will be afraid of making too much money for the principal.[27] There might be a variety of reasons, of course, why such contracts are infeasible or undesirable. If, for example, the principal is uncertain about either the true value of R or the extent of the agent's confirmatory bias q, she may not have enough information to propose a contract that always leads the agent to invest optimally.

Nevertheless, even in the presence of uncertainty the principal would generally be better off if she could restrain the agent's ability to take an extreme action. If feasible to restrain the agent, the principal could propose a contract that stipulates that the agent cannot invest more than a fraction of her wealth in any single asset. Even more simply, the principal could simply give the agent only a portion of her wealth to invest. All of these strategies serve the same purpose, namely preventing an overconfident agent from investing too much of the principal's money in a single asset while still taking advantage of the information that the agent actually does possess.

VI. Discussion and Conclusion

We believe that confirmatory bias is important in many social and economic situations, and that variants of the formulation developed in this paper can be usefully applied in formal economic models. For instance, confirmatory bias is likely to matter when a

---

27. From the principal's point of view, the optimal proportional allocation to a risky investment, a^*, maximizes the objective function $V(a) \equiv \mu^*(\theta,q) u(1 + (R - 1)a) + (1 - \mu^*(\theta,q)) \cdot (1 - a)$. The optimal allocation a^* then satisfies the following necessary and sufficient condition:

$$\geq 0, \quad a^* = 1$$
$$\mu^*(\theta,q)u'(1 + (R - 1)a^*)(R - 1) - (1 - \mu^*(\theta,q))u'(1 - a^*) = 0, \quad a^* \in [0,1]$$
$$\leq 0, \quad a^* = 0.$$

The principal would like to propose a contract specifying that the agent receives an arbitrarily small reward when the principal's gross return is $1 + (R - 1)a^*$, no payoff when the principal's gross return is $1 - a^*$, and a large penalty for any other gross return. Such a contract would punish the agent if he chooses an allocation that is more extreme than the principal desires.

decision-maker must aggregate information from many sources. In a setting where several individuals (nonstrategically) transmit their beliefs to a principal, how should she combine these reports to form her own beliefs? If the principal thought that the agents were Bayesians, then she would be very sensitive to the strength of the agents' beliefs. Suppose, for instance, that the principal knows that all agents receive signals of strength $\theta = .6$. Then if two agents report believing Hypothesis A with probability .6 and one agent reports believing Hypothesis B with probability .77 (meaning he has gotten three more b signals than a signals), the principal should believe in Hypothesis B with probability .6.

What if the principal were aware that agents were subject to confirmatory bias? If confirmatory bias is so severe that only an agent's first signal is very informative, then the principal may wish to discount the *strength* of agents' beliefs and basically aggregate according to a "majority rules" criterion. In the example above, for instance, the principal should perhaps think Hypothesis A is more likely, because two of three agents believe in it. We think this intuition has merit, but it is complicated by the fact that agents who believe relatively weakly in a hypothesis may be more likely to be wrong than right. So, if the principal thought confirmatory bias were severe *and* were very sure that all agents had received lots of information, then in our example she should believe that *all three* agents have provided evidence in favor of Hypothesis B. Hence, she should believe *more* in Hypothesis B than she would if the agents were Bayesian.

We suspect nonetheless that the "majority-rules intuition" is more valid, especially when considering realistic uncertainty by the principal about how many signals each agent has received. If she were highly uncertain about how much information each agent received, she would assume weak beliefs merely reflected that an agent got few signals. Similarly, if the principal thinks susceptibility to confirmatory bias is heterogeneous, she might infer that an agent's weak beliefs indicate merely that he is not susceptible to overconfidence, and count weak beliefs as much as strong beliefs. Indeed, she may then count them *more* heavily, since confirmation-free agents are not only less likely to be overconfident, they are also less likely to be wrong.

This intuition that, when aggregating information from a group, it may be wise to count the number of people with given

beliefs rather than the strength of their convictions suggests a related prescription for organizational design: relative to what she would do with Bayesian agents, a principal may prefer to hire more agents to collect a given amount of information. That is, while the lower value of information processing by confirmatory agents may mean that either more or fewer should be hired than if they were Bayesian, fixing the total amount of information processing a principal wants done, with confirmatory agents she should prefer more people thinking than if they were fully rational agents.

Imagine, for instance, that a principal allocated 1000 "signals" among different agents, whose reports she would aggregate to form her own beliefs. There are various costs that might influence how many agents to have, or (equivalently) how many signals to allocate per agent, e.g., the fixed cost of hiring new agents and decreasing returns from each individual due to fatigue or the increasing opportunity cost of time. But the optimal number of confirmatory agents is likely to be greater than the optimal number of Bayesian agents. Intuitively, the value of allocating a signal to a confirmatory agent is less than the value of allocating it to a Bayesian agent, *unless* the confirmatory agent is unbiased by previous signals. For example, if the principal hires 1000 confirmatory agents, each to report his observation of a single signal, then she receives all of the information contained in the signals. If the principal instead hires one confirmatory agent to report his beliefs after interpreting 1000 signals, she may get far less information. Both signal allocations would yield the same amount of information if the agents were Bayesian.

We suspect that a similar issue plays out less abstractly in different aspects of the legal system. While other explanations are probably more important, confirmatory bias may help to explain some features of the American jury system, such as the bias toward more rather than fewer jurors and the use of a majority-rules criterion with no mechanism (other than jury deliberations) to extract the strength of all participants' convictions. Confirmatory bias may also help to justify the use of multiple judges to reach a decision when using a single judge seems to be more cost-effective. Appeals, for example, are usually heard by a panel of judges that does not include the trial judge, and some legal scholars (e.g., Resnik [1982]) argue that the judge who adjudicates at trial should not also supervise settlement bargaining and

pretrial discovery, which is the process by which litigants request information from each other. These observers fear that the trial judge might learn things during pretrial activities that would "bias" her during the trial. The notion that the quality of the judge's decisions during the trial suffers if she has more information relevant to the case is somewhat puzzling; worries that she can be "biased" by more information certainly flies in the face of the Bayesian model. While there are various types of bias that one could imagine (e.g., that the judge will use her rulings during the trial to punish perceived misbehavior during the discovery process), the evidence on confirmatory bias raises the possibility that the judge will form preconceptions during the discovery phase of litigation that will cause her to misread additional evidence presented at trial.

Finally, the discovery process itself nicely illustrates how the polarization associated with confirmatory bias may have important implications. Discovery takes place when potential litigants think that a trial is relatively likely, and hence wish to engage in the costly effort of preparing for that trial. Nonetheless, litigants often settle their case out of court during or after the discovery process. Discovery encourages this settlement by promoting the exchange of information between the litigants and, hence, helping to align their perceptions of the likely outcome at trial. But while the evidence garnered during the discovery process sometimes does lead to settlement before trial, confirmatory bias suggests that the discovery process may be less efficient at achieving such settlement than would be hoped: if a piece of evidence is ambiguous, it may move the parties' beliefs farther apart. Each litigant will interpret the evidence through the prism of his or her own beliefs, and each may conclude that the evidence supports his or her case. More generally, efforts to reduce disagreements by providing evidence to the parties involved in a conflict may not be as easy to achieve as one would hope.

Much of our discussion above implicitly makes an assumption about judgment whose psychological validity has not (to our knowledge) been determined by research: that somebody designing an institution is aware of the bias of others. We suspect that usefully incorporating confirmatory bias into economic analysis will depend upon the extent to which people believe that others suffer from confirmatory bias. It could be that people are well aware of biases in others' judgment, or that people are unaware of

the general tendency toward confirmatory bias.[28] Investors who hire a money manager might or might not believe that the money manager suffers from a confirmatory bias (and is therefore prone toward overconfidence). A principal hiring an employee to make decisions might or might not know that the employee will be prone to making such errors. By the logic of economic models that involve multiple agents, these distinctions are likely to matter: Just as assuming that rationality is common knowledge is often very different than merely assuming that people are rational, assuming that agents are aware of others' irrationality may be very different than merely assuming that people are irrational.

How might economic implications depend on people's awareness of others' confirmatory bias? One possibility is that people might exploit the bias of others. A principal may, for instance, design an incentive contract for an agent that yields the agent lower wages on average than the agent anticipates, because the agent will be overconfident about her judgments in ways that may lead her to exaggerate her yield from a contract. Conversely, others may wish to mitigate bias rather than exploit it. A principal may be more concerned with overcoming costly bias of an agent than with exploiting it, and design contracts that avoid errors.

APPENDIX 1: DIFFERENTIAL-STRENGTH SIGNALS AND UNDERCONFIDENCE

If the agent receives signals of different strengths in different periods, it is possible that the agent will be *underconfident* in his belief about which of the two states is most likely. Suppose, for example, that the agent receives three signals $s_t \in \{a,b\}$, $t \in \{1,2,3\}$. Suppose that the first two signals are distributed according to prob $(s_t = a|A) =$ prob $(s_t = b|B) = \theta > 0.5$, $t \in \{1,2\}$, but that the agent's third signal is distributed according to prob $(s_3 = a|A) =$ prob $(s_3 = b|B) = \theta^3/[\theta^3 + (1 - \theta)^3]$. That is, the agent's third signal is three times as strong as first- or second-period signals. As before, with probability $q > 0$ the agent misreads signals that conflict with his belief about which state is more likely. (This

---

28. Unfortunately, while this issue may turn out to be central to economic applications of confirmatory bias (and to applications of other psychological biases), we have not found psychological research that convincingly resolves this issue. There is a small literature in "construal" that concerns third-party awareness of biases. See, e.g., Ross [1987], and tangentially Paese and Kinnaly [1993]. We have not found investigation of this issue in the context of confirmatory bias or overconfidence.

means that the probability of misreading is independent of the strength of the signal.)

Suppose that the agent perceives that his first two signals support Hypothesis B, while his third signal supports Hypothesis A. Formally, the agent perceives $(\sigma_1 = \beta, \sigma_2 = \beta, \sigma_3 = \alpha)$. Given these perceived signals, the agent's posterior likelihood ratio is $\Lambda(s_1 = b, s_2 = b, s_3 = a) = \theta/(1 - \theta) > 1$. Now, suppose that a Bayesian observer knows both that the agent's posterior likelihood ratio is $\Lambda = \theta/(1 - \theta)$ and that the agent suffers from confirmatory bias. Given the distributions of the signals, the observer is able to infer that the agent has perceived $(\sigma_1 = \beta, \sigma_2 = \beta, \sigma_3 = \alpha)$. Then, the observer's belief regarding the relative likelihood that the state is $x = A$ versus $x = B$ is given by

$$\Lambda^*(b,\beta,\alpha) = \frac{(1 - \theta)(1 - \theta + q\theta)(1 - q)\theta^3}{\theta(\theta + q(1 - \theta))(1 - q)(1 - \theta)^3}$$

$$= \frac{(1 - \theta + q\theta)\theta^2}{(\theta + q(1 - \theta))(1 - \theta)^2} > \frac{\theta}{1 - \theta}, \qquad \forall q \in (0,1].$$

Therefore, given what she infers about the agent's sequence of perceived signals, a Bayesian observer believes that the biased agent is *underconfident* in his belief that the true state is A.

This underconfidence result arises here because the observer infers the exact sequence of the agent's perceived signals from his likelihood ratio. In this light, the results here are the same as the path-dependent underconfidence example in the text—if the agent is known to have only recently come to believe in a hypothesis, then he will be underconfident. In our main model, in which the agent receives signals of equal strength, an observer who knows the agent's beliefs cannot infer the exact sequence of the agent's perceived signals.

While there may be some domains in which this differential-signal model is applicable, constructing examples of underconfidence seem to require clever contrivance. It is first of all clear that the "overconfidence" result will be stronger than the underconfidence result in one sense: in the model of this paper, the overconfidence result holds for *all* final beliefs by the agent. Any underconfidence example will clearly hold for only *some* final beliefs—because it will always be the case that a confirmatory agent is overconfident when all his perceived signals favor one hypothesis.

We suspect, moreover, that more complicated and weaker versions of Proposition 1 will hold in more general models. The underconfidence result seems to rely on the agent having received a small number of signals, where certain final beliefs can only be generated by a unique path of updating. Consequently, it is very likely that a "limit overconfidence" result would hold—once an agent is likely to have received large numbers of signals of all strengths, we can assure that $\Lambda^* < \Lambda$ when $\Lambda > 1$.

<div align="center">APPENDIX 2: PROOFS</div>

Proof of Proposition 1. We first notice that

$$\text{prob}\,(n_\alpha, n_\beta | A) = \sum_{i=0}^{n_\beta} \text{prob}\,(i, i | A) c(n_\alpha - n_\beta, n_\alpha + n_\beta - 2i)$$

$$\cdot \theta[\theta + q(1 - \theta)]^{n_\alpha - 1 - i}[(1 - q)(1 - \theta)]^{n_\beta - i}$$

and

$$\text{prob}\,(n_\alpha, n_\beta | B) = \sum_{i=0}^{n_\beta} \text{prob}\,(i, i | B) c(n_\alpha - n_\beta, n_\alpha + n_\beta - 2i)$$

$$\cdot (1 - \theta)[(1 - \theta) + q\theta]^{n_\alpha - 1 - i}[(1 - q)\theta]^{n_\beta - i},$$

where $c(n_\alpha - n_\beta, n_\alpha + n_\beta - 2i)$ is the number of ways to choose $n_\alpha - n_\beta$ more a signals than b signals in $n_\alpha + n_\beta - 2i$ draws without ever having chosen an equal number of a and b signals, and prob $(i, i | x)$ is the probability of observing i perceived a and i perceived b signals in $2i$ draws when the true state is $x \in \{A, B\}$. Given the symmetric distribution of the signals, prob $(i, i | A) = $ prob $(i, i | B)$.[29] Therefore, prob $(n_\alpha, n_\beta | A)$ and prob $(n_\alpha, n_\beta | B)$ differ

---

29. Formally,

$$\text{prob}\,(i, i | A) = \text{prob}\,(i, i | B) = \sum_{j=0}^{i} \sum_{k=0}^{i-j} \sum_{i=0}^{\max\{i-j-k-1, 0\}}$$

$$\cdot d_{jkl}\theta^j\theta^{**k}((1 - q)(1 - \theta))^{j+k}(1 - \theta)^{i-j-k-l}\theta^{*l}((1 - q)\theta)^{i-j-k}.$$

The coefficient d_{jkl} is the number of ways to choose j signals in favor of the correct hypothesis when the agent believes the two hypotheses are equally likely, k biased signals favoring the correct hypothesis, $i - j - k - l$ signals in favor of the incorrect hypothesis when the agent believes the two hypotheses are equally likely, l biased signals in favor of the incorrect hypothesis, $j + k$ unbiased signals opposing a belief in favor of the correct hypothesis, and $i - j - k$ unbiased signals opposing a belief in favor of the incorrect hypothesis.

only by the effect of the signals that the agent perceives after the last time that he believes the two hypotheses are equally likely.

Using Bayes' Rule,

$$(1.1) \quad \Lambda^*(n_\alpha, n_\beta) = \frac{\text{prob}\,(n_\alpha, n_\beta | A)}{\text{prob}\,(n_\alpha, n_\beta | B)}$$

$$= \frac{\begin{array}{c}\Sigma_{i=0}^{n_\beta}\,\text{prob}\,(i,i|A)c(n_\alpha - n_\beta, n_\alpha + n_\beta - 2i) \\ \cdot\, \theta[\theta + q(1-\theta)]^{n_\alpha - 1 - i}[(1-q)(1-\theta)]^{n_\beta - i}\end{array}}{\begin{array}{c}\Sigma_{i=0}^{n_\beta}\,\text{prob}\,(i,i|B)c(n_\alpha - n_\beta, n_\alpha + n_\beta - 2i)(1-\theta) \\ \cdot\, [(1-\theta) + q\theta]^{n_\alpha - 1 - i}[(1-q)\theta]^{n_\beta - i}\end{array}}.$$

Because $[\theta + q(1-\theta)]/[(1-\theta) + q\theta] < \theta/(1-\theta)$, $\forall q \in (0,1]$, it follows that

$$(1.2) \quad [\theta + q(1-\theta)]^{n_\alpha - 1 - i}(1-\theta)^{n_\alpha - 1 - i} \le [(1-\theta) + q\theta]^{n_\alpha - 1 - i}\theta^{n_\alpha - 1 - i}$$

with a strict inequality for $i = 0$ since the hypotheses imply that $n_\alpha \ge 2$. Factoring and multiplying (1.2) by $(1-\theta)(1-q)^{n_\beta - i}$ and rearranging, we have

$$(1.3) \quad \theta[\theta + q(1-\theta)]^{n_\alpha - 1 - i}(1-q)^{n_\beta - i}(1-\theta)^{n_\beta - i}$$

$$\le (1-\theta)[(1-\theta) + q\theta]^{n_\alpha - 1 - i}(1-q)^{n_\beta - i}\theta^{n_\beta - i}\left(\frac{\theta}{1-\theta}\right)^{n_\alpha - n_\beta}$$

$\forall i$, with a strict inequality for at least $i = 0$ since $n_\alpha \ge 2$. Using (1.1), (1.3), and prob $(i,i|A) = \text{prob}\,(i,i|B)$,

$$\Lambda^*(n_\alpha, n_\beta) < \frac{\begin{array}{c}\Sigma_{i=0}^{n_\beta}\,\text{prob}\,(i,i|A)c(n_\alpha - n_\beta, n_\alpha + n_\beta - 2i)(1-\theta) \\ \cdot\, [(1-\theta) + q\theta]^{n_\alpha - 1 - i}(1-q)^{n_\beta - i}\theta^{n_\beta - i}(\theta/(1-\theta))^{n_\alpha - n_\beta}\end{array}}{\begin{array}{c}\Sigma_{i=0}^{n_\beta}\,\text{prob}\,(i,i|B)c(n_\alpha - n_\beta, n_\alpha + n_\beta - 2i)(1-\theta) \\ \cdot\, [(1-\theta) + q\theta]^{n_\alpha - 1 - i}(1-q)^{n_\beta - i}\theta^{n_\beta - i}\end{array}}$$

$$= \left(\frac{\theta}{1-\theta}\right)^{n_\alpha - n_\beta} = \Lambda(n_\alpha, n_\beta).$$

Proof of Proposition 2. Clearly $\lim_{\epsilon \to 0} \Lambda^*(n_\alpha, 0 | 1 - \epsilon, 1 - \epsilon) = \infty \forall n_\alpha > 0$ and $\lim_{\epsilon \to 0} \Lambda^*(n_\alpha, 1 | 1 - \epsilon, 1 - \epsilon) = (n_\alpha + 1)/(n_\alpha - 1) \forall \cdot n_\alpha > 1$.

It can be shown that, if the agent's current beliefs are that

A and B are equally likely, and A is true,

then the probability that the next signal is $\alpha \approx 1$ is $\beta = \epsilon$

A and B are equally likely, and B is true,
    then the probability that the next signal is $\alpha = \epsilon$ is $\beta \approx 1$
A is probably true, and A is true,
    then the probability that the next signal is $\alpha \approx 1$ is $\beta = \epsilon^2$
A is probably true, and B is true,
    then the probability that the next signal is $\alpha \approx 1$ is $\beta \approx \epsilon$
B is probably true, and A is true,
    then the probability that the next signal is $\alpha \approx \epsilon$ is $\beta \approx 1$
B is probably true, and B is true,
    then the probability that the next signal is $\alpha = \epsilon^2$ is $\beta \approx 1$.

From these numbers we can calculate that, if $n_\alpha > n_\beta$,

· Suppose that A is the true state. Consider all paths σ^* such that (1) $\sigma_1^* = \beta$ and (2) there is always a strict majority of β signals until $2n_\beta - 1$ signals, after which all signals are α. Then the probability of any particular path σ^* is about $\epsilon^{n_\beta+1}$. All other paths each occur with probability on the order of $\epsilon^{n_\beta+2}$ or greater when $n_\beta \geq 2$.

· Suppose that B is the true state. Consider all paths σ^{**} such that (1) $\sigma_1^{**} = \alpha$ and (2) there is always a strict majority of α signals. The probability of any particular path σ^{**} is about $\epsilon^{n_\beta+1}$. All other paths each occur with probability on the order of $e^{n_\beta+2}$ or greater when $n_\beta \geq 2$.

To show that $\Lambda^*(n_\alpha, n_\beta | 1 - \epsilon, 1 - \epsilon) < 1$ with $n_\beta \geq 2$, therefore, we need only to show that the number of paths of type σ^{**} is strictly greater than the number of paths of type σ^*. This is easy to verify. For every particular path of type σ^*, there exists a path of type σ^{**} that is the mirror image of that path for the first $2n_\beta - 1$ signals (replacing each α with a β and each β with an α), and whose last $n_\alpha - n_\beta + 1$ signals consist of $n_\alpha - n_\beta - 1$ α's followed by 2 β's. In addition, there will exist at least one more path of type σ^{**}; for instance, n_α α's followed by n_β β's.

QED

Proof of Proposition 3. The proof is by induction. Define $\Lambda(n)$ as the agent's relative likelihood ratio after observing n signals. Suppose that $\Lambda(1) > 1$. Then a Bayesian observer infers that the agent observed a single true "a" signal, and $\Lambda^*(1) = \theta/(1 - \theta) > 1$. Now suppose that $\Lambda(n)$, $\Lambda^*(n)$, and $\Lambda(n + 1) > 1$. We must show that $\Lambda^*(n + 1) > 1$. First, suppose that n is an even number. Because $\Lambda(n) > 1$, after period n the agent has perceived at least

two more "a" signals than "b" signals. Therefore, knowing only that $\Lambda(n) > 1$, a Bayesian observer's relative likelihood ratio, $\Lambda^*(n) = \text{prob}\,(x = A)/\text{prob}\,(x = B)$, is given by

$$(3.1)\quad \Lambda^*(n) = \frac{\sum_{j=0}^{(n/2)-1} \sum_{i=0}^{j} p(i,i|A)c(n-2j, n-2i)}{\sum_{j=0}^{(n/2)-1} \sum_{i=0}^{j} p(i,i|B)c(n-2j, n-2i)(1-\theta)} \cdot \frac{\cdot \theta[\theta + q(1-\theta)]^{n-1-j-i}(1-q)^{j-i}(1-\theta)^{j-i}}{\cdot [(1-\theta) + q\theta]^{n-1-j-i}(1-q)^{j-i}\theta^{j-i}} = \frac{p^A(n)}{p^B(n)},$$

where $p(i,i|x)$ and $c(\cdot,\cdot)$ are defined as in the proof of Proposition 1. Define $p^x(n)$ as the probability of perceiving a (strict) majority of "α" signals in n draws given the state $x \in \{A,B\}$. Then $\Lambda^*(n + 1)$ is given by

$$(3.2)\qquad \Lambda^*(n + 1) = \frac{p^A(n) + p(n/2, n/2|A)\theta}{p^B(n) + p(n/2, n/2|B)(1-\theta)}.$$

Because $p(n/2, n/2|A) = p(n/2, n/2|B)$ and $\theta > 0.5$, $\Lambda^*(n + 1) > 1$ follows immediately from the hypothesis that $\Lambda^*(n) > 1$, which implies that $p^A(n) > p^B(n)$.

Now suppose that n is an odd number. Because by hypothesis $\Lambda(n) > 1$, after period n the agent has perceived more "a" than "b" signals. Therefore, knowing only that $\Lambda(n) > 1$, a Bayesian observer's relative likelihood ratio, $\Lambda^*(n) = \text{prob}\,(x = A)/\text{prob} \cdot (x = B)$, is given by

$$(3.3)\quad \Lambda^*(n) = \frac{\sum_{j=0}^{(n-1)/2} \sum_{i=0}^{j} p(i,i|A)c(n-2j, n-2i)}{\sum_{j=0}^{(n-1)/2} \sum_{i=0}^{j} p(i,i|B)c(n-2j, n-2i)(1-\theta)} \cdot \frac{\cdot \theta[\theta + q(1-\theta)]^{n-1-j-i}(1-q)^{j-i}(1-\theta)^{j-i}}{\cdot [(1-\theta) + q\theta]^{n-1-j-i}(1-q)^{j-i}\theta^{j-i}} = \frac{p^A(n)}{p^B(n)}.$$

Meanwhile, $\Lambda^*(n + 1)$ is given by

$$(3.4)\qquad \Lambda^*(n + 1) = \frac{\begin{array}{c} p^A(n) - \sum_{i=0}^{(n-1)/2} p(i,i|A)c(1, n-2i) \\ \cdot \theta[\theta + q(1-\theta)]^{((n-1)/2)-i}[(1-q) \\ \cdot (1-\theta)]^{((n-1)/2)-i}(1-q)(1-\theta) \end{array}}{\begin{array}{c} p^B(n) - \sum_{i=0}^{(n-1)/2} p(i,i|B)c(1, n-2i) \\ \cdot (1-\theta)[(1-\theta) + q\theta]^{((n-1)/2)-i} \\ \cdot [(1-q)\theta]^{((n-1)/2)-i}(1-q)\theta \end{array}}.$$

Because $\Lambda^*(n) > 1$ implies that $p^A(n) > p^B(n)$, in order to establish $\Lambda^*(n + 1) > 1$ it is sufficient to show that

$$(3.5) \quad \sum_{i=0}^{(n-1)/2} p(i,i|A)c(1,n - 2i)\theta[\theta + q(1 - \theta)]^{((n-1)/2)-i}[(1 - q)$$

$$\cdot (1 - \theta)]^{((n-1)/2)-i}(1 - q)(1 - \theta)$$

$$\leq \sum_{i=0}^{(n-1)/2} p(i,i|B)c(1,n - 2i)(1 - \theta)[(1 - \theta)$$

$$+ q\theta]^{((n-1)/2)-i}[(1 - q)\theta]^{((n-1)/2)-i}(1 - q)\theta.$$

Using the fact that $p(i,i,|A) = p(i,i|B)$ and canceling like terms, the inequality in (3.5) is satisfied if

$$[\theta + q(1 - \theta)]^{((n-1)/2)-i}(1 - \theta)^{((n-1)/2)-i}$$

$$\leq [1 - \theta + q\theta]^{((n-i)/2)-1}\theta^{((n-1)/2)-i} \quad \forall i \in \{0, \ldots, (n - 1)/2\}.$$

But this inequality is always satisfied because $(\theta + q(1 - \theta))/((1 - \theta) + q\theta) < \theta/(1 - \theta) \forall q \in (0,1]$. Therefore, $\Lambda^*(n + 1) > 1$.

QED

Proof of Proposition 4. The first hypothesis implies that $\theta^* > 0.5$, and therefore, using Lemma 1, P_W satisfies

$$P_W = (1 - \theta) \cdot [(1 - p(1,\theta^*)) + p(1,\theta^*) \cdot P_W]$$

$$+ \theta \cdot [p(1,\theta^{**}) \cdot P_W],$$

or

$$P_W = \frac{(1 - \theta) \cdot (1 - ((1 - \theta^*)/\theta^*))}{(1 - (1 - \theta) \cdot ((1 - \theta^*)/\theta^*) - \theta((1 - \theta^{**})/\theta^{**}))} \cdot$$

$P_W > 0$ because $\theta^{**} > 0.5$ for all $q \geq 0$.

The second hypothesis implies that $\theta^* \leq 0.5$, and therefore, using Lemma 1, P_W satisfies

$$P_W = (1 - \theta) \cdot P_W + \theta \cdot [p(1,\theta^{**}) \cdot P_W].$$

$P_W = 0$ because $p(1,\theta^{**}) < 1$.

QED

Proof of Proposition 5. Ignoring integer problems, and since $q > 1 - 1/(2\theta)$ for the cases we consider below, the definition of

$D(\mu)$ and Lemma 1 imply that

$$P_W(\mu) = \left(1 - \left[\frac{1 - \theta^*}{\theta^*}\right]^{D(\mu)}\right) + \left[\frac{1 - \theta^*}{\theta^*}\right]^{D(\mu)} P_W(0.5) > 0.$$

(i) Note that $\lim_{q \to 1} \theta^* = 1$ for all (μ,θ). Therefore, for q sufficiently close to 1, $P_W(\mu)$ can be made arbitrarily close to 1.

(ii) Note that $\lim_{\theta \to 0.5} \theta^* > 0.5$ and $\lim_{\theta \to 0.5} D(\mu) = \infty$ for all (μ,q). Therefore, for θ sufficiently close to 0.5, $P_W(\mu)$ can be made arbitrarily close to 1.

$$\text{QED}$$

Proof of Proposition 6. (i) Fix $\theta \in (0.5,\underline{\mu}\,]$. The result follows directly from the fact that the principal's payoff from having the agent observe signals, $\Pi(\theta,q) = \mu^*(\theta,q)u(WR) + (1 - \mu^*(\theta,q))u(0)$, is continuously monotone decreasing in q, with $\Pi(\theta,1 - 1/2\theta) > u(W)$ and $\Pi(\theta,1) \leq u(W)$. (ii) Fix $q \in (0,1]$. The result follows directly from the fact that $\Pi(\theta,q)$ is continuously monotone increasing in θ, with $\Pi(0.5,q) \leq u(W)$ and $\Pi(u,q) > u(W)$.

$$\text{QED}$$

Proof of Proposition 7. The proof is by construction. In order to establish the result, it is sufficient to show that there exists a function $g(\cdot)$ such that

$$\frac{g(u(WR)) - g(u(W))}{g(u(W)) - g(u(0))} \leq \frac{1 - \mu^*(\theta,q)}{\mu^*(\theta,q)}\ .$$

Define the function $g(\cdot)$ as

$$g(x) = \begin{cases} x, x \leq u(W) \\ u(W) + \epsilon(x - u(W)), x > u(W). \end{cases}$$

Clearly, $g' > 0, g'' \leq 0$, and

$$\frac{g(u(WR)) - g(u(W))}{g(u(W)) - g(u(0))} = \frac{\epsilon(u(WR) - u(W))}{u(W) - u(0)} \leq \frac{1 - \mu^*(\theta, q)}{\mu^*(\theta, q)}$$

for ϵ sufficiently small. Parts (ii) and (iii) follow directly from concavity of $g(\cdot)$ and Theorem 1 in Pratt [1964].

$$\text{QED}$$

UNIVERSITY OF CALIFORNIA, BERKELEY
EMORY UNIVERSITY

REFERENCES

Arkes, H. R., "Principles in Judgment/Decision Making Research Pertinent to Legal Proceedings," *Behavioral Sciences and the Law,* VII (1989), 429–456.

Bacon, F. (1620), *The New Organon and Related Writings* (New York, NY: Liberal Arts Press, 1960).

Baumann, A. O., R. B. Deber, and G. G. Thompson, "Overconfidence among Physicians and Nurses: The 'Micro-Certainty, Macro-Uncertainty,' Phenomenon," *Social Science and Medicine,* XXXII (1991), 167–174.

Beattie, J., and J. Baron, "Confirmation and Matching Biases in Hypothesis Testing," *Quarterly Journal of Experimental Psychology: Human Experimental Psychology,* XL (1988), 269–297.

Bjorkman, M., "Internal Cue Theory: Calibration and Resolution of Confidence in General Knowledge," *Organizational Behavior and Human Decision Processes,* LVIII (1994), 386–405.

Bodenhausen, G. V., and M. Lichtenstein, "Social Stereotypes and Information-Processing Strategies: The Impact of Task Complexity," *Journal of Personality and Social Psychology,* LII (1987), 871–880.

Bodenhausen, G. V., and R. S. Wyer, "Effects of Stereotypes in Decision Making and Information-Processing Strategies," *Journal of Personality and Social Psychology,* XLII (1985), 267–282.

Borum, R., R. Otto, and S. Golding, "Improving Clinical Judgment and Decision Making in Forensic Evaluation," *Journal of Psychiatry and Law,* XXI (1993), 35–76.

Bruner, J., and M. Potter, "Inference in Visual Recognition," *Science,* CXLIV (1964), 424–425.

Camerer, C., "Individual Decision Making," in *Handbook of Experimental Economics,* J. Kagel and A. E. Roth, eds. (Princeton, NJ: Princeton University Press, 1995), pp. 587–703.

Chapman, L. J., and J. P. Chapman, "Genesis of Popular but Erroneous Psychodiagnostic Observations," *Journal of Abnormal Psychology,* LXXII (1967), 193–204.

Chapman, L. J., and J. P. Chapman, "Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs," *Journal of Abnormal Psychology,* LXXIV (1969), 271–280.

Chapman, L. J., and J. P. Chapman, "Test Results Are What You Think They Are," *Psychology Today,* V (1971), 106.

Darley, J., and P. Gross, "A Hypothesis-Confirming Bias in Labeling Effects," *Journal of Personality and Social Psychology,* XLIV (1983), 20–33.

Devine, P. G., E. R. Hirt, and E. M. Gehrke, "Diagnostic and Confirmation Strategies in Trait Hypothesis Testing," *Journal of Personality and Social Psychology,* LVIII (1990), 952–963.

Dougherty, T. W., D. B. Turban, and J. C. Callender, "Confirming First Impressions in the Employment Interview: A Field Study of Interviewer Behavior," *Journal of Applied Psychology,* LXXIX (1994), 659–665.

Einhorn, H., and R. Hogarth, "Confidence in Judgment: Persistence of the Illusion of Validity," *Psychological Review,* LXXXV (1978), 395–416.

Feller, W., *An Introduction to Probability Theory and Its Applications* (New York, NY: John Wiley and Sons, Inc., 1968).

Fischoff, B., and R. Beyth-Marom, "Hypothesis Evaluation from a Bayesian Perspective," *Psychological Review,* XC (1983), 239–260.

Fischhoff, B., P. Slovic, and S. Lichtenstein, "Knowing with Certainty: The Appropriateness of Extreme Confidence," *Journal of Experimental Psychology: Human Perception and Performance,* III (1977), 552–564.

Fleming, J., and A. J. Arrowood, "Information Processing and the Perseverance of Discredited Self-Perceptions," *Personality and Social Psychology Bulletin,* V (1979), 201–205.

Friedrich, J., "Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena," *Psychological Review,* C (1993), 298–319.

Griffin, D., and A. Tversky, "The Weighing of Evidence and the Determinants of Confidence," *Cognitive Psychology,* XXIV (1992), 411–435.

Hamilton, D. L., and T. L. Rose, "Illusory Correlation and the Maintenance of Stereotypic Beliefs," *Journal of Personality and Social Psychology,* XXXIX (1980), 832–845.

Hamilton, D. L., S. J. Sherman, and C. M. Ruvolo, "Stereotype-Based Expectancies: Effects on Information Processing and Social Behavior," *Journal of Social Issues,* XLVI (1990), 35–60.

Haverkamp, B. E., "Confirmatory Bias in Hypothesis Testing for Client-Identified and Counselor Self-Generated Hypotheses," *Journal of Counseling Psychology,* XL (1993), 303–315.

Hodgins, H. S., and M. Zuckerman, "Beyond Selecting Information: Biases in Spontaneous Questions and Resultant Conclusions," *Journal of Experimental Social Psychology,* XXIX (1993), 387–407.

Hubbard, M., "Impression Perseverance: Support for a Cognitive Explanation," *Representative Research in Social Psychology,* XIV (1984), 48–55.

Jennings, D. L., M. R. Lepper, and L. Ross, "Persistence of Impressions of Personal Persuasiveness: Perseverance of Erroneous Self-Assessments outside the Debriefing Paradigm," *Personality and Social Psychology Bulletin,* VII (1981), 257–263.

Jennings, D. L., T. M. Amabile, and L. Ross, "Informal Covariation Assessment: Data-Based versus Theory-Based Judgments," in *Judgment under Uncertainty: Heuristics and Biases,* D. Kahneman, P. Slovic, and A. Tversky, eds. (Cambridge: Cambridge University Press, 1982), pp. 211–230.

Keren, G., "Facing Uncertainty in the Game of Bridge: A Calibration Study," *Organizational Behavior and Human Decision Processes,* XXXIX (1987), 98–114.

———, "On the Ability of Monitoring Non-Veridical Perceptions and Uncertain Knowledge: Some Calibration Studies," *Acta Psychologica,* LXVII (1988), 95–119.

Klayman, J., and Y.-w. Ha, "Confirmation, Disconfirmation, and Information in Hypothesis Testing," *Psychological Review,* XCIV (1987), 211–228.

Lepper, M. R., L. Ross, and R. R. Lau, "Persistence of Inaccurate Beliefs about the Self: Perseverance Effects in the Classroom," *Journal of Personality and Social Psychology,* L (1986), 482–491.

Lord, C. G., L. Ross, and M. R. Lepper, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology,* XXXVII (1979), 2098–2109.

Macan, T. H., and R. L. Dipboye, "The Effects of the Application on Processing of Information from the Employment Interview," *Journal of Applied Social Psychology,* XXIV (1994), 1291–1314.

Mahajan, J., "The Overconfidence Effect in Marketing Management Predictions," *Journal of Marketing Research,* XXIX (1992), 329–342.

Mahoney, M., "Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System," *Cognitive Therapy and Research,* I (1977), 161–175.

Mehle, T., C. F. Gettys, C. Manning, S. Baca, and S. Fisher, "The Availability Explanation of Excessive Plausibility Assessments," *Acta Psychologica,* XLIX (1981), 127–140.

Miller, A. G., J. W. McHoskey, C. M. Bane, and T. G. Dowd, "The Attitude Polarization Phenomenon: Role of Response Measure, Attitude Extremity, and Behavioral Consequences of Reported Attitude Change," *Journal of Personality and Social Psychology,* LXIV (1993), 561–574.

Nisbett, R., and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, NJ: Prentice-Hall, 1980).

O'Donoghue, E., and M. Rabin, "Incentives for Procrastinators," Northwestern University, CMSEMS Discussion Paper No. 1181, February 25, 1997.

Oskamp, S., "Overconfidence in Case-Study Judgments," in *Judgment under Uncertainty: Heuristics and Biases,* D. Kahneman, P. Slovic, and A. Tversky, eds. (Cambridge: Cambridge University Press, 1982), pp. 287–293.

Paese, P. W., and M. Kinnaly, "Peer Input and Revised Judgment: Exploring the Effects of (Un)Biased Confidence," *Journal of Applied Social Psychology,* XXIII (1993), 1989–2011.

Perkins, D. N., *The Mind's Best Work* (Cambridge, MA: Harvard University Press, 1981).
Pfeifer, P. E., "Are We Overconfident in the Belief That Probability Forecasters Are Overconfident?" *Organizational Behavior and Human Decision Processes,* LVIII (1994), 203–213.
Plous, S., "Biases in the Assimilation of Technological Breakdowns: Do Accidents Make Us Safer?" *Journal of Applied Social Psychology,* XXI (1991), 1058–1082.
Popper, Karl, *Conjectures and Refutations* (London: Routledge and Kegan Paul, Ltd., 1963).
Pratt, J., "Risk Aversion in the Small and in the Large," *Econometrica,* XXXII (1964), 122–136.
Redelmeier, D., and A. Tversky, "On the Belief That Arthritis Pain is Related to the Weather," *Proceedings of the National Academy of Sciences USA,* XCIII (1996), 2895–2896.
Resnik, J., "Managerial Judges," *Harvard Law Review,* XCVI (1982), 374–445.
Ross, L., "The Problem of Construal in Social Inference and Social Psychology," in *A Distinctive Approach to Psychological Research: The Influence of Stanley Schachter,* R. E. N. Neil, E. Grunberg, Judith Rodin, and Jerome E. Singer, eds. (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1987), pp. 118–150.
Soll, J., "Determinants of Overconfidence and Miscalibration: The Roles of Random Error and Ecological Structure," *Organizational Behavior and Human Decision Processes,* LXV (1996), 117–137.
Souter, G., "Overconfidence Causes Reinsurers to Underprice Coverages: Consultant," *Business Insurance,* XXVII (1993), 47.
Stangor, C., "Stereotype Accessibility and Information Processing," *Personality and Social Psychology Bulletin,* XIV (1988), 694–708.
Stangor, C., and D. N. Ruble, "Strength of Expectancies and Memory for Social Information: What We Remember Depends on How Much We Know," *Journal of Experimental Social Psychology,* XXV (1989), 18–35.
Tomassini, L. A., I. Solomon, M. B. Romney, and J. L. Krogstad, "Calibration of Auditors' Probabilistic Judgments: Some Empirical Evidence," *Organizational Behavior and Human Performance,* XXX (1982), 391–406.
Van Lenthe, J., "ELI: An Interactive Elicitation Technique for Subjective Probability Distributions," *Organizational Behavior and Human Decision Processes,* LV (1993), 379–413.
Winman, A., and P. Juslin, "Calibration of Sensory and Cognitive Judgments: Two Different Accounts," *Scandinavian Journal of Psychology,* XXXIV (1993), 135–148.
Wood, A. S., "Fatal Attractions for Money Managers," *Financial Analysts Journal,* XLV (1989), 3–5.
Wyatt, D., and D. Campbell, "On the Liability or Stereotype of Hypothesis," *Journal of Abnormal and Social Psychology,* XLVI (1951), 496–500.
Zuckerman, M., C. R. Knee, H. S. Hodgins, and K. Miyake, "Hypothesis Confirmation: The Joint Effect of Positive Test Strategy and Acquiescence Response Set," *Journal of Personality and Social Psychology,* LXVIII (1995), 52–60.

Perceptions of Others' Political Affiliation Are Moderated by Individual Perceivers' Own Political Attitudes

John Paul Wilson, Nicholas O. Rule*

Department of Psychology, University of Toronto, Toronto, Ontario, Canada

Abstract

Previous research has shown that perceivers can accurately extract information about perceptually ambiguous group memberships from facial information alone. For example, people demonstrate above-chance accuracy in categorizing political ideology from faces. Further, they ascribe particular personality traits to faces according to political party (e.g., Republicans are dominant and mature, Democrats are likeable and trustworthy). Here, we report three studies that replicated and extended these effects. In Study 1a, we provide evidence that, in addition to showing accuracy in categorization, politically-conservative participants expressed a bias toward categorizing targets as outgroup members. In Study 1b, we replicate this relationship with a larger sample and a stimulus set consisting of faces of professional politicians. In Study 2, we find that trait ascriptions based on target political affiliation are moderated by perceiver political ideology. Specifically, although Democrats are stereotyped as more likeable and trustworthy, conservative participants rated faces that were categorized as Republicans in Study 1a as more likeable and trustworthy than faces categorized as Democrats. Thus, this paper joins a growing literature showing that it is critical to consider perceiver identity in examining perceptions of identities and traits from faces.

Introduction

People can be quite accurate at extracting a number of seemingly concealable social identities from facial information alone. Unlike categorizations based on race, sex, and age, other group memberships (e.g., sexual orientation, religion, and political affiliation) are quite perceptually ambiguous. Nevertheless, perceivers reliably exhibit above-chance accuracy for distinguishing members of these groups [1]. These categorizations, at least for some dimensions, seem to be driven by the perception of subtle cues that may differ between groups [2]. Accordingly, much of the existing research on accuracy in categorizing ambiguous group members has focused on the target, leaving much unknown about the perceiver's contribution to these judgments.

For example, studies have found that participants can accurately distinguish political affiliation based on photos of faces [3], [4], [5], [6] [7], [8]. Furthermore, Rule and Ambady [4] found that these effects seem to have been driven by traits attributed to the faces; specifically, power (a composite of ratings of dominance and facial maturity) and warmth (a composite of ratings of likeability and trustworthiness). Republican faces were perceived as more powerful than Democrats and, to the extent that a face was perceived as powerful, it was more likely to be categorized as Republican. On the other hand, the warmer a face was perceived, the more likely it was to be categorized as a Democrat. Samochowiec et al. [6] reported similar findings: perceived dominance partially accounted for the relationship between targets' perceived and actual political ideology in Europe.

Other studies have also found relationships between facial traits and perceived political ideology. For example, Berggren, Jordahl, and Poutvaara [9] found that conservative politicians in Finland were more attractive than candidates on the political left. This is similar to a result reported by Bull and Hawkes [10] in which politicians judged to be conservative were more attractive, intelligent-looking, and of higher social class than those judged to be more liberal. These judgments can have electoral consequences. Though Bull et al. [11] found that these ratings did not correspond to vote share, Olivola et al. [8] found that politicians in conservative geographic areas tend to benefit in terms of electoral outcomes if they have a conservative-looking appearance.

Although the extant research has provided important information about factors that may underlie categorizations of faces according to political party affiliation, it may be limited in some critical ways. For example, the effects described by Rule and Ambady [4] result from a focus primarily on the target in isolation from the perceiver. It is important to note that this research draws from ideas based in an ecological theory of perception [12] adapted to theories of ecological social perception [13], [14]. These theories argue that faces signal certain things to perceivers about what the target may afford. A target may appear, for instance, more or less trustworthy or dominant [15], [16], which may lead the perceiver to trust or fear that person, accordingly. However, it is likely that the accuracy of categorizing ambiguous targets is driven at least in part by the perceivers' identities, dispositions, or states. For example, it is known that perceptions of

ambiguous group members may be influenced by perceivers' attitudes toward the groups [17] or exposure to members of the group [18]. Brambilla et al. [18] found that heterosexual perceivers who had more experience with gay men were better at categorizing gay faces. Similarly, it is plausible that categorizations and trait ascriptions of Democrat and Republican faces may differ based on the political ideology of the perceiver/judge. There does exist some evidence that perceiver identities and ideologies influence categorizations and judgments based on political affiliation. Jahoda [5], for instance, found that people categorized faces as political ingroup or outgroup members largely as a function of likeability. More recently, Chiao et al. [19] found that perceiver gender influences ratings towards male and female politicians.

Target political affiliation may interact with perceiver identity in another important way. In addition to accurate perceptions of political ideology, Samochowiec et al. [6] also found that participants were more likely to classify faces as outgroup members than as ingroup members. These findings are consistent with a more general ingroup overexclusion effect [20], [21]. The ingroup overexclusion effect is thought to be a result of motivated social cognitions related to social identity, such that people tend to be protective of the ingroup. As a result of this protectiveness, perceivers may show a default bias toward categorizing others as outgroup members. This may be the case especially when groups are perceptually ambiguous. For example, Blascovich, Wyer, Swart, and Kibler [22] observed an ingroup overexclusion effect for racially ambiguous targets, and Castano et al. [21] found that Northern Italians (who had strongly identified as such) were more likely to exclude ambiguous targets that had a mix of Northern and Southern Italian features.

The current work represents another attempt to incorporate perceiver identities into understanding the legibility of target political affiliation from faces. The work conducted by Rule and Ambady [4], for example, was limited in that it employed a sample that was demographically liberal in an area where Democratic politicians tend to be highly favored by the public at large. Findings suggesting that Republicans are perceived as more powerful and Democrats as more warm may be the result of a biased sample. In fact, Olivola et al. [8] failed to find a relationship between perceived political party and judgments of traits such as honesty and dependability with an Internet sample from the US. Aggregate perceptions across diverse sets of perceivers may not correspond with target political party, highlighting the importance of the possible role that individual differences in perceiver identities may exert upon judgments. The current research therefore included participants across the political spectrum from an Internet community sample. This made it possible to investigate how categorizations and trait ascriptions might differ based on perceivers' political orientation.

In three studies, we investigated how perceivers' political ideology influenced their perceptions of faces from different political parties. In each study, rather than use a single-item scale or self-categorization of party membership, we asked participants to complete a validated scale measuring political values: the McClosky political conservatism scale [23], [24]. This allowed us to more precisely measure a continuous relationship between political ideology and perceptions of targets according to perceived party. We used the revised version of the scale reported by McClosky and Bann [24], including both the 19-item Classical Conservatism scale and seven items addressing social welfare issues, for a total of 26 items. This 26-item version has been used as a broad measure of conservatism-liberalism [25], [26]. For each

item on the scale, participants can select either the conservative response or the liberal response (or neither).

In Study 1, we tested the relationship between perceivers' political leanings and their categorizations of targets as Democrats and Republicans. In addition to expecting to replicate past results showing above-chance accuracy overall, we expected perceiver ideology to influence categorizations. Based on prior research on social groups and categorization [20], we predicted that participants would tend to *overexclude* targets from the ingroup. In other words, we predicted that conservatives would show a default tendency to categorize faces as Democrat outgroup members, and that liberals would show a default tendency to categorize faces as Republican outgroup members. Thus, we expected individuals' political beliefs to bias their categorizations in a way that guards their ranks against potential adulteration by outgroup members, consistent with theories of ingroup overexclusion [20].

Further, and more important, in Study 2 we extended beyond previous work to predict that perceivers' political leanings would moderate how personality traits would be differentially ascribed to Democrat and Republican faces. As past work sampling participants from a highly politically liberal population showed a relationship between Democrats and warmth, we therefore tested whether this is a function of the targets (i.e., warmer faces are universally more likely to be seen as Democrats) or of the perceivers (i.e., warmer faces are more likely to be seen as ingroup members for both conservatives and liberals). The latter result would be consistent with previous reports showing that voters tend to associate likeability and warmth with their favored political candidates, which tend to be those who share their political beliefs [5], [27], [28]. In other words, we hypothesized that conservative participants would not show the previously described tendency to rate Democrat faces as more likeable and trustworthy. Rather, these participants were expected to see *Republican* faces as more likeable and trustworthy—thereby showing evidence of ingroup favoritism [29]. Moreover, we predicted that these relationships would be mediated by the extent to which a face was likely to be categorized as Republican. Thus, conservative perceivers should see faces that belong to actual Republicans as likeable to the extent that they can be accurately categorized by party affiliation.

Study 1a

The present study aimed to expand on previous work showing that individuals are able to accurately perceive others' political affiliation [3], [4], [5], [6], [7], [8]. This work has largely but not exclusively (see [5], [8], [19]) focused on the targets of perception, rather than the characteristics of the perceivers making the judgments. Some of these studies asked participants to self-report their affiliation with particular political parties but reported little or no relationship between the participants' political party and their categorizations of the faces [3], [4], [5]. Rather than inquire about party membership, one study asked participants to self-report their political leaning along a continuous scale ranging from conservative to liberal [6].

The present study aimed to expand on this past work by assessing participants' political attitudes with the McClosky political conservatism scale [23], [24]. Like Samochowiec et al. [6], we expected that participants would show an ingroup overexclusion effect whereby participants with higher scores on the conservatism scale would be more likely to categorize targets as Democrats versus Republicans (and, as such, participants on the liberal end of the scale would show the opposite tendency). Although Democrats and Republicans are not isomorphic with liberals and conservatives, the two correspond tightly; see

Abramowitz and Saunders (2008) for discussion and Study 1a results for data illustrating this. In addition, the current design uses discrete categorizations of targets into specific political groups (Democrats and Republicans), as opposed to the scalar measure of political behavior used in Samochowiec et al.'s [6] study. Although the method used by Samochowiec et al. [6] allows for greater sensitivity in assessing accuracy, a potential advantage of the dichotomous categorization task here is that it may allow for a more precise estimate of participants' response bias measure than was possible in the preceding work, which required bifurcation of the scale to somewhat artificially divide the targets into two groups.

Method

Ethics statement. This study was approved by the Research Ethics Board at the University of Toronto. All participants provided written informed consent prior to participation in the study.

Participants. Forty-six American participants engaged in the experiment via Amazon's Mechanical Turk [30]; two did not complete the experiment and their data were removed from analysis for a total of 44 participants [23 male (52%), M_{Age} = 32 years, SD = 11; 35 White (80%), 3 Hispanic (7%), 2 Black (5%), 1 Asian (2%), 3 Multiracial/Other (7%)]. No other participants were removed from analysis for any other reason. The self-reported party affiliation of our participants was as follows: 8 Republican (18%); 11 Democrat (25%); 7 Libertarian (16%); 1 Green (2%); 14 Independent (32%); 3 other (7%).

Stimuli. Stimuli were borrowed from a previous study examining judgments of Americans' political party affiliation [4]. Photos consisted of 60 male and female Democrat (n = 30) and Republican (n = 30) university students from a small liberal arts college in the northeast US. Among Democrats, 15 were male and 15 were female. Among Republicans, 21 were male and 9 were female. All were Caucasian. All of the photos were digitally scanned from the portrait section of the students' senior yearbooks. The students had indicated either the Democrat or Republican student organization among their extra-curricular activities in the yearbook; however, the photos were not taken from club photos or other contexts in which the students' political affiliations would have been salient at the time of photography. Each of the images was cropped to the extremes of the head: top of hair, bottom of chin, and sides of ears. All of the images were grayscale and standardized to be of equal face height while maintaining the aspect ratio of the face. The stimuli are available upon request.

Procedure. Participants were instructed that they would be seeing a series of faces presented on their computer's screen and that they were to categorize them as Democrats or Republicans as quickly and accurately as possible, relying on their "gut" instinct. All of the face stimuli described above were presented in random order and participants provided their responses via mouse-click. After categorizing all of the faces, participants were asked to complete the McClosky political conservatism scale and some demographic questions, including the opportunity to self-categorize among several conservative (Republican, Libertarian, Tea), liberal (Democrat, Green), or neither (independent, other) parties. Consistent with reports of the polarization of political parties in the contemporary US, despite diverse self-categorizations outside of the traditional bipartisan split [31], participants' conservatism scores were highly correlated with their categorical party affiliations for the subsets identifying as Democrats or Republicans and more broadly as any of the liberal or conservative parties: all r's>.71.

Results

Data from Study 1a are available here http://dx.doi.org/10.6084/m9.figshare.967938. Participants' categorizations of the faces as Democrat and Republican were analyzed using the signal detection theory statistics A' and B'' to measure sensitivity and response bias, respectively (see Macmillan & Creelman [32] for formulas). A' is a nonparametric measure of recognition accuracy that does not require homogeneous variance and can be calculated when participants exhibit hit or false alarm rates of 1 or 0. B'' is a complementary measure of response bias. Correct categorizations of Democrats were counted as hits and incorrect categorizations of Republicans were counted as false-alarms. Because we coded correct categorizations of Democrats as hits, negative B'' values represent a tendency to categorize targets as Democrats, whereas positive values represent a tendency to categorize targets as Republicans.

Replicating previous work with an undergraduate laboratory sample [4], participants were significantly more accurate than chance in categorizing the targets' political affiliations: $M_{A'}$ = .63, SD = .12, $t(43)$ = 7.15, $p<.001$, Cohen's d = 1.07. Overall, participants showed no significant bias toward categorizing targets as Democrats or Republicans ($M_{B''}$ = −.03, SD = .15), $t(43)$ = 1.49, $p = .14$, Cohen's $d = 0.22$.

Critically, participants' response bias was significantly correlated with their levels of political conservatism. Following the analyses of Peterson-Badali et al. [25], we calculated the percentage of conservative responses chosen by participants on the McClosky scale: $M = 39\%$, SD = 22%; inter-item reliability Cronbach's $\alpha = .86$. It should be noted that this scale also provides an index of liberalism but, as these were strongly negatively correlated with the percentage of conservative responses ($r = −.77$), all results in the present work are discussed in terms of conservatism. Looking at the relationship between ideology and response bias (B''), we observed a greater tendency for participants to categorize targets as Democrats as a function of their endorsement of conservative items on the scale, suggesting an ingroup-overexclusion effect [14]: $r(42)$ = −.29, $p = .052$—though this correlation did not reach conventional levels of statistical significance (i.e., $p<.05$). Conservatism was not correlated with participants' accuracy (A'): [$r(42)$ = −.14, $p = .36$], conceptually replicating past work in which the political affiliation of perceivers did not correlate with accuracy in categorizing US politicians [33] (those authors also examined response bias, but did not report as to whether it was associated with political affiliation).

Study 1b

Study 1b was designed to replicate Study 1a while addressing a few possible shortcomings. First, the relationship between conservatism and response bias did not reach conventional levels of significance, perhaps due to issues with statistical power. Here, we therefore increased the sample size in order to investigate whether the ingroup overexclusion effect is reliable. Next, Study 1a used a limited stimulus set of non-politicians. Here, we used a larger set of faces, this time of professional American politicians, in order to establish the generalizability of the observed relationships. Finally, faces were not blocked according to target sex in Study 1a, above. As female politicians may be more likely to be Democrats [34], presenting the faces in uniform blocks by target sex might reduce participants' ability to rely on sex to infer political party. Thus, Study 1b used a design consisting of blocks by target sex.

Method

Ethics statement.   This study was approved by the Research Ethics Board at the University of Toronto. All participants provided written informed consent prior to participation in the study.

Participants.   One hundred twenty-one American participants engaged in the experiment via Amazon's Mechanical Turk [30] [56 male (46%), M_{Age} = 35 years, SD = 12; 87 White (72%), 7 Hispanic (6%), 8 Black (7%), 10 Asian (8%), 9 Multiracial/Other (7%)]. No participants were removed from analysis for any reason. The self-reported party affiliation of our participants was as follows: 21 Republican (17%); 57 Democrat (47%); 6 Libertarian (5%); 4 Green (3%); 16 Independent (13%); 17 other (14%).

Stimuli.   Stimuli were once again borrowed from a previous study examining judgments of Americans' political party affiliation [4]. Here, however, we used photos of professional politicians who ran for the US Senate in 2004 and 2006. Each photo was cropped to the extremes of the targets' heads (top of head, bottom of chin, sides of hair or ears), converted to grayscale, and standardized for size. Ethnic minority targets were not used in this study. Daniel Akaka, a non-White Democrat, was mistakenly included in the stimulus set, but analyses excluded him. In total, we used 115 target faces: 58 Democrats (n = 15 women) and 57 Republicans (n = 5 women). All stimuli are available as supplemental files here http://dx.doi.org/10.6084/m9.figshare.967942.

Procedure.   The main procedure was identical to Study 1a with the exception that the faces were presented randomly within randomly-ordered blocks that were uniform in gender. After completing the face categorization task and the McClosky scale, participants were asked to indicate their political party identification in an open-ended format. We also asked participants to list any faces that they had recognized. We performed our analyses on the complete dataset, as well as after removing individual trials in which a participant recognized a face (0.68% of all trials, in total), and the results did not change appreciably; we therefore report the full data below.

Results

Data from Study 1b are available here http://dx.doi.org/10.6084/m9.figshare.967939. As in Study 1a, participants' categorizations of the faces as Democrat and Republican were analyzed using A' and B'' to measure sensitivity and response bias, respectively. Correct categorizations of Democrats were counted as hits and incorrect categorizations of Republicans were counted as false-alarms. Replicating Study 1a, participants were significantly more accurate than chance in categorizing the targets' political affiliations: $M_A{}' $ =.54, SD =.09, $t(120)$ = 5.12, $p<.001$, Cohen's $d = 0.47$. Unlike Study 1a, participants showed a slight bias towards categorizing targets as Republicans ($M_B{}''$ =.009, SD =.04), $t(120)$ = 2.39, p = .02, Cohen's $d=0.19$.

As in Study 1a, we calculated the percentage of conservative responses chosen by participants on the McClosky scale: $M = 33\%$, SD = 21%; inter-item reliability Cronbach's α = .88. Participants in Study 1b were slightly less conservative than in Study 1a. Looking at the relationship between ideology and response bias (B''), we saw that participants once again showed a greater tendency to categorize targets as Democrats as a function of their endorsement of conservative items on the scale: $r(119)$ = −.24, p = .007. As in Study 1a, conservatism was not correlated with participants' accuracy (A'): $r(119)$ = −.12, p = .18.

Discussion

As predicted, participants in both Studies 1a and 1b showed above-chance accuracy in categorizing target faces according to political party affiliation. This replicates past work on the accuracy of judging political orientation from static facial information alone and, notably, does so for target samples of both professional politicians and non-politicians. Thus, party membership is expressed through appearance generally for both politicians and non-politicians, as shown in previous work [4]. Further, we also found that participants showed a tendency to overcategorize targets as outgroup members in both studies. This reinforces previous work showing a similar relationship between estimated response bias and participant political affiliation using targets in European parliaments [6]. Thus, as social identity theory researchers have posited, and as our results confirm, people tend to set a relatively high threshold for categorizing an ambiguous target as an ingroup member. This is sensible, as one way to maintain a positive social identity is to be cautious about whom one allows into the ingroup [20], [21].

This result additionally suggests that the encoding of ambiguous social category information may be influenced by social contextual factors [21], [35]. That is, faces transmit a wealth of information, much of which is accurately identified by perceivers. However, this process is influenced not just by target-based affordances, but also by the motivations of the perceiver. Study 2 explored how perceiver identities may also influence the relationship between perceived political party and perceived personality traits.

Study 2

The results of Studies 1a and 1b showed that perceivers' method of categorizing targets as Democrats and Republicans was influenced by their personal political beliefs. As individuals endorsed more conservative values, they were significantly less likely to think that targets were also conservatives, suggesting an ingroup overexclusion effect [20]. Previous research on perceptions of Democrats and Republicans reported that members of the two groups were associated with different personality traits [4]. Specifically, Republicans were perceived to be significantly more dominant and facially mature than were Democrats and this difference partially accounted for accuracy in judging the targets' political affiliation. Interestingly, that work also found that— irrespective of actual political group membership—targets consensually believed to be Democrats were perceived as high in likeability and trustworthiness, whereas targets consensually believed to be Republicans were perceived as high in dominance and facial maturity. Given that the participants in these studies were undergraduates at a relatively politically liberal university in the Boston metropolitan area, we wondered whether the relationship between perceived political party and perceived personality traits might be moderated by individual variation in political beliefs. Thus, in Study 2, we asked an online community sample of participants to assess the personality traits of the same targets and related their judgments to a measure of their political attitudes. We predicted that participants would rate Republican faces as more likeable and trustworthy to the extent that they report more conservative attitudes.

Method

Ethics statement.   This study was approved by the Research Ethics Board at the University of Toronto. All participants provided written informed consent prior to participation in the study.

Participants.   Fifty-nine American participants engaged in the experiment via Amazon's Mechanical Turk; twelve did not complete the experiment and their data were removed from analysis for a total of 47 participants [21 male (45%), M_{Age} = 35

years, *SD* = 13; 37 White (79%), 1 Hispanic (2%), 5 Black (11%), 4 Asian (9%); 19 Democrat (40%), 5 Republican (11%), 3 Green Party (6%), 2 Libertarian (4%), 13 Independent (28%), 5 Other(11%)].

Stimuli. Stimuli were the same as in Study 1a.

Procedure. Participants were instructed that they would be seeing a series of faces presented on their computer's screen and that they were to rate them on various traits as quickly and accurately as possible, relying on their "gut" instinct. Participants viewed blocks of faces organized by trait such that they saw each face in random order within each block and rated it on one of either dominance (rated from 1 = "Submissive" to 7 = "Dominant"), facial maturity (rated from 1 = "Babyish" to 7 = "Mature"), likeability (rated from 1 = "Not at all likeable" to 7 = "Very likeable"), or trustworthiness (rated from 1 = "Not at all trustworthy" to 7 = "Very trustworthy") before moving on to the next trait block; participants provided their responses via mouse-click and all of the target faces were presented in each block. Trait-blocks were presented in random order. Once participants had completed each of the four blocks, they were asked to complete the 26-item McClosky political conservatism scale and some demographic questions. No mention of politics or political parties was made at any point during study recruitment or execution; thus, participants were ostensibly rating faces on personality traits and nothing more.

Results

Data from Study 2 are available here http://dx.doi.org/10.6084/m9.figshare.967940.

Replication of past work. As in the previous studies, the sample was more inclined to endorse liberal versus conservative values: on average, participants endorsed only 32% (*SD* = 18%) of the conservative statements on the McClosky scale (inter-item reliability Cronbach's α = .77). Similar to previous work [4], which analyzed the data with targets rather than participants as the unit of analysis (here, all inter-rater reliabilities Cronbach's α's>.92), Republican targets were perceived as significantly more dominant [*r*(58) = .34, *p* = .007] and facially mature [*r*(58) = .28, *p* = .03] than were Democrat targets but showed no differences for likeability [*r*(58) = −.14, *p* = .29] or trustworthiness [*r*(58) = −.17, *p* = .19]. The lack of a significant relationship between targets' party affiliation and ratings of their likeability and trustworthiness is inconsistent with some past research [4], though the correlation coefficients are in the expected direction. This lack of a statistically significant correlation may reflect the fact that this sample was likely more conservative than some samples used in past work. In addition, the extent to which targets were perceived as looking Republican (based on the categorizations made in Study 1a) was positively associated with how dominant [*r*(58) = .40, *p* = .002] and facially mature [*r*(58) = .31, *p* = .02] the targets were rated, and was negatively associated with how trustworthy [*r*(58) = −.25, *p* = .05] and likeable [*r*(58) = −.23, *p* = .08] they were rated, though the latter correlations were only marginally significant. Thus, the present data replicated the relationships reported in previous research [4].

Main analysis: Perceiver political beliefs and trait ratings. Of more pertinence to the central question, however, accounting for individual differences in participants' political beliefs substantially moderated these relationships. As we were primarily interested in the relationship between participants' political beliefs and their perceptions of the targets according to political party, we conducted the analyses with the participants as the unit of analysis. Hence, we correlated each participant's ratings on each trait with a dummy-coded vector in which Democrat

targets were coded as 0 and Republican targets were coded as 1 and then calculated point-biserial sensitivity correlations indexing the extent to which Republican targets were rated higher on each trait by each participant. In other words, we correlated a target's party identification with trait ratings for each participant. After transforming the resultant correlation coefficients into Fisher's *z* scores, we then correlated them with the participants' endorsement of conservative beliefs measured by the McClosky scale; the resulting value therefore represented the correlation between each participant's individual level of conservatism and the extent to which the participant rated Democrat and Republican targets differently for each trait. Some participants provided the same rating for all of the faces in a particular block/trait, rendering it impossible to calculate a sensitivity correlation; hence, the degrees of freedom for the correlations between the McClosky conservatism scores and face ratings varied slightly between traits.

The more conservative the participants rated themselves to be, the more likely they were to see Republican targets as both more likeable [*r*(44) = .32, *p* = .03] and more trustworthy [*r*(43) = .34, *p* = .02] than Democrat targets. No significant correlations were observed for dominance [*r*(44) = −.07, *p* = .62] and facial maturity [*r*(45) = −.15, *p* = .31]. This suggests that individuals view others who share their political views as more likeable and more trustworthy, which may reflect ingroup favoritism. The null correlation for maturity should perhaps be treated with some caution, as the sample is somewhat small. However, we had no prior reason to predict a relationship between perceived maturity and ingroup affiliation.

To better understand the differences in ratings made to Democrat and Republican targets along these traits and to test the ingroup favoritism hypothesis, we measured the extent to which the perceived political affiliation of the targets might statistically mediate the relationship between actual political affiliation and participants' ratings of the targets' traits as a function of the participants' conservatism. To achieve this, we calculated partial sensitivity correlations for each participant that controlled for the consensus perception of each target's political affiliation, based on the ratings given in Study 1a. We therefore computed the proportion of participants in Study 1a who categorized each target as a Republican, producing a decimal value ranging between 0 and 1 that indexed how "Republican" each target face looked (also used in the replication analyses reported above). We then calculated point-biserial correlations between the targets' actual party membership and the participants' ratings of their likeability and trustworthiness, respectively, while controlling for the degree to which each target looked Republican. We then transformed these partial point-biserial correlation coefficients into Fisher's *z* scores and correlated them with the participants' scores on the McClosky conservatism scale.

After controlling for the extent to which each target looked Republican, the relationship between participants' conservatism and the degree to which they viewed Republicans as more trustworthy than Democrats was reduced in size and no longer significant: *r*(43) = .20, *p* = .19. Similarly, the relationship between participants' conservatism and the degree to which they viewed Republicans as more likeable than Democrats was also reduced in size and yielded a correlation that was only marginally significant after controlling for the extent to which each target looked Republican: *r*(44) = .29, *p* = .054. These findings suggest that more conservative participants' distinction between conservative versus liberal others as more likeable and more trustworthy may rely, in part, on the legibility of the targets' political group membership. Thus, targets who look more Republican and less Democrat may

be liked and trusted more by more conservative participants (see also [8]).

Discussion

As we predicted, Study 2 showed that perceiver political ideology moderates the relationship between perceptions of target party affiliation and trait ascriptions. These results are somewhat consistent with past research in which perceptions of political party membership occurred largely as a function of participants' own political leanings and the favorability of targets based on appearance [5], [8]. We were able to extend upon this past work by statistically demonstrating that favorable personality ratings (such as trustworthiness and likeability) may have been ascribed to the extent that a target was more likely to be categorized by his or her political affiliation. In other words, the legibility of targets' political affiliation may drive favorable ratings for ingroup faces, though it may have done so without perceivers' explicit awareness.

Importantly, these results were obtained by examining the relationship between political categorizations made by one sample of participants and ratings made by a separate sample. As such, participants are not merely ascribing favorable traits to targets that they have personally identified as ingroup members or expressing some autocorrelative effect. Past work in this domain, on the other hand, has more directly linked particular perceivers' political attitudes with the same perceivers' judgments of faces [5] (but see [8]). Furthermore, we found it intriguing that target political affiliation plays a role in trait judgments even when perceivers have not been oriented to the political nature of the task (i.e., they do not realize that these faces have been identified as affiliating with one or another political party). These results are similar to research by Olivola et al. [8] in which candidates who look Republican are more likely to receive the votes of Republican voters. Here, however, we have used a continuous measure of political ideology, so our results are not exclusive to conservative participants – rather, the more conservative one is, the more likeable and trustworthy Republican faces look; and the less conservative one is, the more likeable and trustworthy Democrat faces look. More centrally, this study serves to qualify existing research on perceived differences between Republican and Democrat faces by showing that these differences are moderated by perceivers' ideology. What is perceived as likeable and trustworthy, then, is a function of the interplay between target and perceiver.

General Discussion

Perceptions of political group membership appear to be influenced by perceivers' own political leanings. First, we replicated past research demonstrating that people can accurately categorize others by political affiliation [3], [4], [5], [6] [7], [8]. Second, we observed a propensity for categorizing faces as outgroup members, consistent with the ingroup overexclusion effect [6], [20]. Finally, we found that perceivers' ascriptions of traits to faces differing in political affiliation are driven by the extent to which they can glean political group membership from faces. This provides an important qualification to previous research finding that Republicans are seen as more dominant and less likeable than Democrats [4]. Here, we found that Republicans may actually be liked and trusted *more* by perceivers, but only to the extent that the perceiver is conservative and the face can be correctly categorized as a Republican. Our results imply that previous work suggesting a more general relationship between party affiliation and facial characteristics was not fully justified [4]. Rather, the current results suggest a more nuanced

relationship between target characteristics, perceiver characteristics, and facial judgments.

This work joins other recent papers in finding that that the basic tendency toward accuracy in categorization and stability in trait inferences for members of ambiguous groups is malleable to the social context [18], [36], [37], [38], [39], [40]. In light of this previous research, this is quite sensible. We found that perceivers ascribed favorable trait ratings to targets who were apparently ingroup members. This is perhaps not surprising, as people tend to show ingroup favoritism across a host of domains, a phenomenon that extends to face perception [41]. For example, other research has shown that racially ambiguous faces are more likely to be categorized as outgroup members than ingroup members when the faces display anger [42]. Further, using data-driven reverse correlation methods, Dotsch, Wigboldus, Langner, and van Knippenberg [43] found that people's mental representations of ethnic outgroup faces are biased to be more criminal and less trustworthy than ethnic ingroup faces and, in related research, participants over-allocated faces appearing criminal-like to an outgroup ethnic category [44]. Such phenomena play out in the political domain as well. In one demonstration, participants rated the representativeness of photographs of biracial political candidates, including Barack Obama, which had been either lightened, darkened, or were un-altered [45]. These researchers found that representativeness ratings tracked with political party affiliation – those who shared the candidate's political party rated the lightened image as most representative, whereas those identifying with an opposing party rated the darkened image as most representative. Importantly, these results remained when controlling for racial attitudes.

In light of such work, it is not altogether surprising that trait ascriptions of faces according to political affiliation are malleable. That is, although targets were seen as more dominant and mature (as well as less likeable and trustworthy) as a function of being categorized as Republicans, the relationship between actual target political affiliation and trait ascriptions was only significant for dominance and maturity, and not likeability or trustworthiness. Further, the relationship between target political categorization and trait ratings was moderated by perceiver ideology. Put more simply, more conservative participants judged faces that appear to be Republicans as likeable and trustworthy, just as previous samples of more liberal participants judged faces that appear to be Democrats as likeable and trustworthy [4].

These results would likely have implications for voting behavior. Olivola et al. [8] found that right-leaning voters favored Republican-looking targets, but that left-leaning voters did not favor Democrat-looking targets. In addition, they found that traits similar to likeability and trustworthiness (i.e., honesty and dependability) did not correlate with the likelihood that a target was categorized as a Republican among a sample of professional politicians. However, they did not report whether this correlation might differ based on perceiver ideology and it is therefore not yet specifically known whether Republican voters are responsive to likeability and trustworthiness in Republican faces versus other dimensions that have been shown to predict voting behavior, such as competence [46]. Perhaps a limitation of our work is that we only used actual target faces with existing political identities. In the future, a more complete investigation into the research question should include the type of data-driven approach described above. That is, using reverse correlation methods, would conservatives and liberals differ on how they visualize Republican and Democrat faces? The current work would support such a prediction. Another limitation is that we were unable to use targets reporting varying levels of conservative and liberal ideology on a continuous scale.

Due to the unobtrusive nature in which the targets' photos were acquired, we were only able to gather information regarding political party affiliation. We hope that future research may use non-politician targets with known political ideology as a continuous measure.

It should also be noted that individuals' own political leanings only moderated perceptions of Republican and Democrat faces on certain dimensions. For instance, the correlations between conservatism and perceptions of Republican versus Democrat faces on dominance or maturity were not statistically significant. Rather, Republican faces were rated higher on these traits, unqualified by perceiver ideology. This suggests that there remain some dimensions upon which Republicans and Democrats are widely agreed to differ. It is perhaps sensible that we see moderation by perceiver ideology for traits (i.e., likeability, trustworthiness) that are more clearly positively-valenced whereas less clearly valenced traits (e.g., dominance, facial maturity) may elicit more stable impressions independent of perceiver attitudes. Future work should assess the role of perceiver ideology in

assessment of other facial traits, such as competence, attractiveness, and femininity/masculinity. This could help to clarify which types of traits are most malleable to perceiver identities.

The current research advances understanding of the encoding and interpretation of ambiguous social identities. It demonstrates that perceiver identities and motivations are a critical component of social ecological perception [14]. People give off social perceptual affordances that drive perceiver reactions and judgments. Critically, here we see that these affordances are encoded in ways that serve the motivations of the perceiver. Although people can sometimes accurately extract information about others' social identities, such processes may also be systematically biased in self-serving ways.

Author Contributions

Conceived and designed the experiments: JPW NOR. Performed the experiments: JPW NOR. Analyzed the data: JPW NOR. Wrote the paper: JPW NOR.

References

1. Tskhay KO, Rule NO (2013) Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. Pers Soc Psychol Rev 17: 72–86.
2. Freeman JB, Johnson KL, Ambady N, Rule NO (2010) Sexual orientation perception involves gendered facial cues. Pers Soc Psychol Bull 36: 1318–1331.
3. Olivola CY, Todorov A (2010) Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. Jô Exp Soc Psychol 46: 315–324.
4. Rule NO, Ambady N (2010) Democrats and Republicans can be differentiated from their faces. PLoS ONE 5: 1–7.
5. Jahoda G (1954) Political attitudes and judgments of other people. Jô Abnorm Soc Psychol 49: 330–334.
6. Samochowiec J, Wänke M, Fiedler K (2010) Political ideology at face value. Soc Psychol Personal Sci 1: 206–213.
7. Carpinella CM, Johnson KL (2013). Politics of the face: The role of sex-typicality in trait assessments of politicians. Soc Cogn 31: 770–779.
8. Olivola CY, Sussman AB, Tsetso K, Kang OE, Todorov A (2012). Republicans prefer Republican-looking leaders: Political facial stereotypes predict candidate electoral success among right-leaning voters. Soc Psychol Personal Sci 3: 605–613.
9. Berggren N, Jordahl H, Poutvaara P (2010). The looks of a winner: Beauty and electoral success. Jô Public Econ 94: 8–15.
10. Bull R, Hawkes C (1982). Judging politicians by their faces. Polit Stud 30: 95–101.
11. Bull R, Jenkins M, Stevens J (1983). Evaluations of politicians' faces. Polit Psychol 4: 713–716.
12. Gibson JJ (1979) The ecological approach to visual perception. Boston: Houghton-Mifflin.
13. McArthur LZ, Baron RM (1983) Toward an ecological theory of social perception. Psychol Rev 90:215–238.
14. Zebrowitz LA, Collins MA (1997) Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. Pers Soc Psychol Rev 1: 204–223.
15. Oosterhof NN, Todorov A (2009) Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. Emotion 9: 128–133.
16. Stirrat M, Perrett DI (2010) Valid facial cues to cooperation and trust: Male facial width and trustworthiness. Psychol Sci 21: 349–354.
17. Hugenberg K, Bodenhausen GV (2004) Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. Psychol Sci 15: 342–345.
18. Brambilla M, Riva P, Rule NO (2013) Familiarity increases the accuracy of categorizing male sexual orientation. Pers Individ Diff 55: 193–195.
19. Chiao JY, Bowman NE, Gill H (2008). The political gender gap: Gender bias in facial inferences that predict voting behavior. PLOS ONE 3:e3666.
20. Leyens JP, Yzerbyt VY (1992) The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. Eurô Jô Soc Psychol 22: 549–569.
21. Castano E, Yzerbyt V, Bourguignon D, Seron E (2002). Who may enter? The impact of in-group identification on ingroup/out-group categorization. Jô Exp Soc Psychol 38:315–322.
22. Blascovich J, Wyer NA, Swart LA, Kibler JL (1997). Racism and racial categorization. Jô Pers Soc Psychol 72: 1364–1372.
23. McClosky H (1958) Conservatism and personality. The Am Polit Sci Rev 52:27–45.
24. McClosky H, Bann CA (1979) On the reappraisal of the classical conservatism scale. Polit Methodol 6: 149–172.
25. Peterson-Badali M, Morine SL, Ruck MD, Slonim N (2004) Predictors of maternal and early adolescent attitudes toward children's nurturance and self-determination rights. Jô of Early Adolesc 24:159–179.
26. Knight K (1999). Liberalism and conservatism. In: Robinson JP, Shaver PR, Wrightsman L, editors. Measures of Political Attitudes. San Diego: Academic Press.
27. Lau RR (1982) Negativity in political perception. Polit Behav 4: 353–377.
28. Markus GB (1982) Political attitudes during an election year: A report on the 1980 NES panel study. Am Polit Sci Rev 76: 538–560.
29. Brewer MB (1999) The psychology of prejudice: Ingroup love or outgroup hate. Jô Soc Issues 55: 429–444.
30. Buhrmeister M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 6:3–5.
31. Abramowitz AI, Saunders KL (2008) Is polarization a myth? Jô Polit 70: 542–555.
32. Macmillan NA, Creelman CD (2005). Signal detection theory and psychophysics. New York: Wiley.
33. Olivola CY, Todorov A (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. Jô Exp Soc Psychol 46: 315–324.
34. Matland RE, King DC (2002). Women as candidates in congressional elections. In: Rosenthal CS, editor. Women transforming Congress. Norman: University of Oklahoma Press.
35. Miller SL, Maner JK, Becker DV (2010) Self-protective biases in group categorization: Threat cues shape the psychological boundary between "us" and "them." Jô Pers Soc Psychol 99: 62–77.
36. Rule NO, Garrett JV, Ambady N (2010). Faces and places: Geographic environment influences the ingroup memory advantage. Jô Pers Soc Psychol 98: 343–355.
37. Remedios JD, Chasteen AL, Rule NO, Plaks JE (2011). Impressions at the intersection of ambiguous and obvious social categories: Does gay + black = likable? Jô Exp Soc Psychol 47: 1312–1315.
38. Pauker K, Ambady N, Freeman JB (2013). The power of identity to motivate face memory in biracial individuals. Soc Cogn 31: 780–791.
39. Stern C, West TV, Jost JT, Rule NO (2013) The politics of gaydar: Ideological differences in the use of gendered cues in categorizing sexual orientation. Jô Pers Soc Psychol 104: 520–541.
40. Stern C, West TV, Schoenthaler A (2013). The dynamic relationship between accuracy and bias in social perception research. Soc Personal Psychol Compass 7: 303–314.
41. Hugenberg K, Wilson JP (2013) Faces are central to social cognition. In: Carlston D, editor. The Oxford Handbook of Social Cognition. Oxford: Oxford University Press.
42. Dunham Y (2011) An angry = outgroup effect. Jô Exp Soc Psychol 47: 668–671.
43. Dotsch R, Wigboldus DH, Langner O, van Knippenberg A (2008) Ethnic out-group faces are biased in the prejudiced mind. Psychol Sci 19: 978–980.
44. Dotsch R, Wigboldus DH, van Knippenberg A (2011) Biased allocation of faces to social categories. Jô Pers Soc Psychol 100: 999–1014.
45. Caruso EM, Mead NL, Balcetis E (2009) Political partisanship influences perception of biracial candidates' skin tone. Proc Natl Acad Sci 196: 20168–20173.
46. Olivola CY, Todorov A (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. Jô Nonverbal Behav 34:83–110.

2018

Living with No: Political Polarization and Transformative Dialogue

Erik Cleven

Robert A. Baruch Bush

Judith A. Saul

Living with No: Political Polarization and Transformative Dialogue

Erik Cleven*, Robert A. Baruch Bush**, and Judith A. Saul***

I. INTRODUCTION

Political polarization is a fact in the United States and has been for some time.[1] Voters are more ideological and political parties are less open to compromise and bipartisanship. This has led to government shutdowns and legislative stalemate in Washington, and increasing difficulty for people to have productive conversations about politics. Political communication on television and online is less civil than face to face communication.[2] In many cases people withdraw from political participation as a result of polarization, exercise self-censorship and experience stress and even trauma. Political polarization also narrows individuals' understanding of themselves and their relationships with others. It makes it difficult to live with difference and diversity.[3]

This gives rise to the question of whether dispute resolution processes can help address the challenges political polarization raise. We argue that dispute resolution processes should not be seen as a substitute for the political process, but rather a complement that can help strengthen it. Based on this view, and on the authors' experience with dialogue work in the former Yugoslavia, as well as in urban and rural settings in the United States, we argue that transformative processes, specifically an approach we call *Transformative Dialogue*, are best suited to addressing the challenges of political polarization both in the United States and internationally. This is because the primary goal of transformative processes is not to reach agreement or find common ground, but rather to change the quality of conflict interactions from negative and destructive to positive and constructive. Transformative dialogue is about helping people gain their voice and choose identities and interactions that otherwise would be closed to them. It does so by supporting participants

---

* Erik Cleven, PhD, is Associate Professor, Department of Politics, Saint Anselm College, Manchester, NH; Board Member, Institute for the Study of Conflict Transformation (ISCT).
** Robert A. Baruch Bush, is Rains Distinguished Professor of Law, Maurice A. Dean School of Law, Hofstra University, Hempstead, NY; Founding Fellow and Board Member, ISCT.
*** Judith A. Saul, is Former Director, Community Dispute Resolution Center, Ithaca, NY; Certified Transformative Mediator; Fellow and Board Member, ISCT.

1. Many, but not all scholars would support the claim that America is politically polarized. For an alternative view *see* Morris P. Fiorina, Samuel A. Abrams, and Jeremy C. Pope, *Polarization in the American Public: Misconceptions and Misreading*, 70 J. POL. 556 (2008). For an overview of the debate between these scholars and critics, see *id*. at 556. Several of those claiming that America is polarized also make the point that this polarization has a long history, so if the question is whether polarization is new, there might be more agreement. *See, e.g.*, JAMES E. CAMPBELL, POLARIZED: MAKING SENSE OF A DIVIDED AMERICA (2016); ANDREW HARTMAN, A WAR FOR THE SOUL OF AMERICA: A HISTORY OF THE CULTURE WARS (2015).

2. DIANA C. MUTZ, IN-YOUR-FACE POLITICS: THE CONSEQUENCES OF UNCIVIL MEDIA 2 (2016).

3. In many other parts of the world, ethnic divisions also polarize politics, with conflict not just over political ideology, but even basic constitutional questions and therefore control of the state. DONALD L. HOROWITZ, ETHNIC GROUPS IN CONFLICT 187-88 (1985).

in gaining clarity about themselves and their interaction with others, and in considering the perspectives of others while maintaining their otherness.

II. POLITICAL POLARIZATION

The polarization of American politics is often identified by commentators as a serious challenge for American democracy.[4] In fact, politics in America have been polarized for some time. This is the case for voters, politicians, and political parties. There are significant differences among voters, with fewer moderates overall as well as fewer liberals in the Republican Party and fewer conservatives in the Democratic Party.[5] Ratings of Republican and Democratic politicians by interest groups like the NRA or Planned Parenthood overlap less and less over time.[6] These developments have taken place at the same time that membership in civic associations has declined, accompanied by a decrease in generalized trust,[7] while online communication is increasing. When people talk face to face they follow norms of politeness and civility; online communication not only facilitates uncivil political discourse, but gives those discourses a large audience.[8] The term "culture wars" also signals that views on cultural issues increasingly correlate with political identity.[9] This makes communication more challenging because there is less agreement about basic principles or outcomes. The other side becomes more "other," more alien, and understanding the other side is harder to imagine.[10]

Political polarization has serious consequences which challenge the functioning of democracies. In a polarized environment, people self-censor themselves and participate less because they fear the reactions of others who hold different opinions than their own.[11] The natural tendency of individuals to socialize with others like themselves is strengthened. While this can have positive effects, like promoting participation because of the in-group trust generated, it also lowers tolerance of others because individuals are not exposed to others with differing viewpoints.[12] These tendencies reinforce and lead to more polarization. After the 2016 elections many

---

4. Matthew Levendusky and Neil Malhotra, *The Media Makes Us Think We Are More Polarized Than We Really Are*, POLITICAL POLARIZATION IN AMERICAN POLITICS 106 (John Sides & Daniel J. Hopkins, eds., 2015).

5. CAMPBELL, *supra* note 1, at 2.

6. NOLAN M. MCCARTY, KEITH T. POOLE, & HOWARD ROSENTHAL, POLARIZED AMERICA: THE DANCE OF IDEOLOGY AND UNEQUAL RICHES 5 (2016).

7 *See* ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 49 (2000).

8. MUTZ, *supra* note 2.

9. HARTMAN, *supra* note 1, at 15.

10. In the United States this polarization exists against a backdrop of overall consensus on constitutional issues. Political conflict centers on hot issues like abortion, same-sex marriage, and the role of government in people's lives. In other countries conflict is about constitutional issues and the very nature of the state. In Kosovo, for example, Serbs are resisting inclusion into what is fundamentally an Albanian state. *See* Oison Tansey, *Kosovo: Independence and Tutelage,* J. DEMOCRACY 153, 154 (2009). Minorities in ethnically divided countries face the constant threat of a "tyranny of the majority". *See* ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 250 (1988). In some cases, fear of this tyranny can lead to violence and ethnic cleansing.

11. *See* Andrew F. Hayes, Dietram A. Scheufele, & Michael E. Huge, *Nonparticipation as Self-Censorship: Publicly Observable Political Activity In A Polarized Opinion Climate,* 28 POL. BEHAVIOR 259 (2006).

12. *See* DIANA CAROLE MUTZ, HEARING THE OTHER SIDE: DELIBERATIVE VERSES PARTICIPATORY DEMOCRACY 77 (2006).

people expressed strong emotional reactions including uneasiness, fear, and anger.[13] Indeed, CNN ran a headline using the term "post-election stress disorder."[14]

Political polarization and the state of American politics today clearly have serious consequences and there is reason to be concerned about how this will affect American democracy in the long run. Lowered participation and trust, self-censorship, the erosion of civility in political discourse and stalemate in Washington not only weaken the current functioning of democracy, but also ultimately can weaken citizens' very belief in democracy as a legitimate political system.

All this raises the question of whether dispute resolution processes can contribute to addressing these serious challenges. We argue that the right kind of dispute resolution process can do so. It must give people a voice and allow them to choose how to understand themselves and their relation to others and, especially, to live with difference. To do this, a dispute resolution process must not promise to "solve" problems or focus on "getting to yes." Rather it must allow people to disagree while still acknowledging the fundamental humanity of those on the other side. If the process can create more positive conflict interactions, people can disagree yet still live and work together. In short, these processes can help address political polarization if, rather than focusing on "getting to yes," they help us to learn to "live with no."[15]

We argue that transformative approaches to conflict,[16] particularly transformative dialogue,[17] are best suited to addressing these challenges because they do not seek to establish common ground but instead focus on the quality of conflict interaction, whatever the outcome of the process and the frequency or infrequency of future interaction.

---

13. A Pew Research Center survey found that while 51% of voters were hopeful after Trump's victory, 53% felt uneasy, 41% sad, and 31% angry. Pew Research Center, *Low Marks for Major Players in 2016 Election —Including the Winner*, Nov. 2016, at 2. When broken down by vote choice it is clear that among Clinton voters 91% felt uneasy (for Trump voters the figure is just 13%), 76% felt scared (5% for Trump voters) and only 7% felt hopeful (96% for Trump voters). *Id.* at 11. Clinton voters used words like shocked, disappointed and disgusted most often to describe how they felt after Trump's election with words like horrified, sad and devastated also frequently mentioned. *Id.* at 7. Nonetheless, 60% of Republicans expressed a wish for the GOP to be more conservative and 49% of Democrats expressed a wish for the Democratic Party to be more liberal, results suggesting that polarization may get worse, not better. *Id.* at 24. The authors' experience working with dialogue confirms this tendency toward differing perceptions of political reality and lack of desire to understand the other. When we have told some people that we are facilitating dialogue between liberals and conservatives, some people have expressed an aversion to the very idea of speaking to someone from "the other side." Some people express fatigue and exhaustion at having to constantly think about politics and race relations. However, we also find people longing for meaningful conversation and listening across the political divide. This attests to people's desire for moral connection to others. *See infra* notes 18-22, 38 and accompanying text.

14. Jenny Gold, *'Post-election stress disorder' strikes on both sides*, CNN (Feb. 20, 2017, 6:23 AM), http://www.cnn.com/2017/02/20/health/post-election-stress-partner/index.html.

15. ROGER FISHER, WILLIAM URY, & BRUCE PATTON, GETTING TO YES: NEGOTIATING AGREEMENT WITHOUT GIVING IN (1991). In this classic book, the authors argue that principled or "win-win" bargaining is the best approach to disputes of all kinds, and helps parties "get to yes". The authors of this article argue that in a polarized polity, "living with no" is a valid goal and important achievement.

16. *See* ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: THE TRANSFORMATIVE APPROACH TO CONFLICT (2nd Ed., 2005).

17. *See* ERIK CLEVEN, WHO NEEDS TO TALK TO WHOM ABOUT WHAT AND HOW?: TRANSFORMATIVE DIALOGUE IN SETTINGS OF ETHNOPOLITICAL CONFLICT (2011).

III. TRANSFORMATIVE DIALOGUE

A. The Transformative Approach to Conflict

The foundation for the transformative approach to conflict was laid with the publication of the first edition of *The Promise of Mediation* in 1994.[18] The book challenged the outcome-oriented nature of standard mediation processes and argued that even though many mediators tried to let solutions come from disputing parties themselves, the pressures to show the success of mediation, and the measurement of success primarily by the number of agreements reached, meant that in reality most mediation was outcome-oriented and mediator driven rather than party driven.[19] The transformative approach was based on several core premises that led to the unique nature of transformative practice.

The first of these premises is that human identity is defined by a balance between concern for self and concern for others. Human beings have a need for self-expression and independent agency. At the same time, individuals value connection to and understanding of others.[20] When conflicts occur this balance is disturbed and human identity is threatened; the transformative approach therefore claims that conflict is best understood as a crisis in human interaction.[21]

In addition to this, the transformative approach is based on the idea that people have the inherent *capacity* for self-determined choice and responsiveness to others.[22] However, the experience of conflict diminishes this capacity. Therefore, third party processes that focus on empowerment -- gaining more clarity about one's situation and the choices one faces – and recognition -- the ability to take the perspective of others -- are best suited to helping individuals regain the capacity for choice and responsiveness to others.

Operating from these premises means that transformative processes are genuinely party driven.[23] The key question an intervener first asks of the parties is, "[w]ho needs to talk to whom, about what, and how?"[24] This question acknowledges that people themselves are best positioned to decide who needs to be part of a conversation, what the conversation needs to be about and how they can best have that conversation. By contrast, most mediation approaches are outcome driven and

---

18. ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: RESPONDING TO CONFLICT THROUGH EMPOWERMENT AND RECOGNITION (1994).

19. *See* Robert A. Baruch Bush, *Staying in Orbit or Breaking Free: The Relationship of Mediation to the Courts Over Four Decades*, 84 N.D. L. REV. 705, 727-32, 735-38 (2008); Robert A. Baruch Bush & Joseph P. Folger, *Mediation and Social Justice: Risks and Opportunities,* 27 OHIO STATE J. ON DISP. RESOL. 1, 22-28 (2012) (both articles referencing and summarizing research that documents the tendencies stated in the text).

20. BUSH & FOLGER, *supra* note 16, at 59-62 (arguing that this view of human identity is reflective of a "relational worldview" emerging in many fields and disciplines).

21. *Id.* at 45-62 (contrasting the transformative view with rights-based and problem-solving views of conflict).

22. *Id.* at 54-58.

23. *See* Bush and Folger, *supra* note 19, at 37-45 (describing the thoroughly party-driven character of transformative intervention practices).

24. *See* CLEVEN, *supra* note 17.

Living with No

founded either on a human needs approach[25] or a narrative approach.[26] Processes based on both of these approaches tend to control and limit party interaction, even if they are intended to liberate participants from "zero-sum" biases and oppressive "dominant discourses."[27]

B. Approaches to Multiparty Dialogue

Most dialogue models claim to be less outcome driven than standard mediation and are defined as more open ended. For example, one definition of dialogue states that it is "a communication process that aims to build relationships between people as they share experiences, ideas, and information about a common concern."[28]

Saunders defines dialogue as

…a process of genuine *inter*action through which human beings listen to each other deeply enough to be changed by what they learn. Each makes a serious effort to take others' concerns into her or his own picture, even when disagreement persists. No participant gives up her or his identity, but each recognizes enough of the others' valid human claims that he or she will act differently toward the other.[29] [italics in original]

The problem with this definition is that it requires that participants listen to each other deeply and presumes that acting differently towards one another after the dialogue must be an outcome. In both of the definitions of dialogue cited above, the kind of speech allowed by the definition is limited. This requires that facilitators control the kind of speech taking place, and this can inhibit transformation because conflict is not fully expressed, and what is difficult is not confronted.[30]

---

25. *See, e.g.*, JOHN WEAR BURTON, CONFLICT: RESOLUTION AND PREVENTION (1993); FISHER, URY & PATTON, *supra* note 15.

26. *See, e.g.*, JOHN WINSLADE & GERALD MONK. NARRATIVE MEDIATION: A NEW APPROACH TO CONFLICT RESOLUTION (2001); Sara Cobb, *Empowerment and Mediation: A Narrative Perspective*, 9 NEG. J. 245 (1993).

27. In human needs approaches the third party translates what parties are saying into needs and interests language, which at best implicitly signals that the words the parties have chosen are inadequate or inappropriate, and at worst is patronizing, shutting people down unless they adopt the mediator's language. This is done in the interest of finding agreement. In narrative approaches, mediators "deconstruct" party narratives, only to reconstruct them in ways that the mediator has deemed is liberating and free of oppressive dominant discourses.

28. LISA SCHIRCH AND DAVID W. CAMPT, THE LITTLE BOOK OF DIALOGUE FOR DIFFICULT SUBJECTS: A PRACTICAL, HANDS-ON GUIDE 6 (2007).

29. HAROLD H. SAUNDERS, A PUBLIC PEACE PROCESS: SUSTAINED DIALOGUE TO TRANSFORM RACIAL AND ETHNIC CONFLICTS 82 (2001). One reason that even Saunder's "*sustained dialogue*" is defined this way is that it, like most dialogue models, is based on ideas that resemble Martin Buber's idea of the I-Thou relationship. In an I-Thou relationship people do not see one another as separate, individual entities, but rather as a new whole. A more useful approach to dialogue could be based on the work of philosopher Emmanuel Levinas, who argues that the other person is radically Other and transcendent. But rather than seeing this difference as negative, Levinas shows that it is also the basis of freedom and moral choice. *See* EMMANUEL LEVINAS, TOTALITY AND INFINITY: AN ESSAY ON EXTERIORITY (1969) *[trans. Alphonso Lingis]*.

30. There is an argument to be made for limiting speech in dialogue. Mutz argues that cross-cutting interactions can lead to violence and that a certain level of civility or politeness is needed to reap the benefits of dialogue with those of differing views. *See* MUTZ, *supra* note 12, at 62. However, Transformative Dialogue is a facilitated process, where participants are not left to themselves but supported by a facilitator who makes it possible for strong confrontations to be potentially beneficial.

We suggest that dialogue can be successful even when no direct communication occurs – i.e., if third parties only speak to potential participants one-on-one, or if dialogue only takes place with members of one group, i.e. *intra*group dialogue, rather than between members of different groups (*inter*group dialogue).[31] We suggest further that the role of dialogue facilitators is not to direct the kind of interaction allowed, defining it from the start as listening to one another or acting differently to one another, nor is their role to insist on particular forms of speech – all of which are controlling, outcome-oriented practices. Rather, the intervener's role is to support parties in reclaiming their capacity for moral choice, in deciding on how to see themselves and on how and whether to recognize the other's perspective. This approach is what we call transformative dialogue.[32] We turn now to how this can be done in practice and how it can address the challenges of political polarization we outlined above.

C. Transformative Dialogue

Transformative dialogue is a process in which a third party works with different members of a community in conflict to change the quality of interactions between them in such a way as to increase the amount of pro-social interaction. The process may include individuals as members of groups, social networks, organizations or institutions. Pro-social interaction occurs when, regardless of how often or seldom they interact, parties act from a position of clarity and strength and are open and responsive to others whether they agree or have deep disagreements.[33]

Transformative dialogue is a multiparty process, like other dialogue approaches, and in multiparty processes identity plays a central and complex role.[34] People participate both as individuals and as members of groups. But individuals have many identities and these can be religious, ethnic, political, or social. A person may be African-American, Jewish, Republican, and a father all at once. Each individual has a unique understanding of their identity which is connected to their unique history and life experience, but people also have elements of their identities

---

31. It is important to understand that the value of dialogue can be realized whether or not people meet across group boundaries (intergroup dialogue). Even intragroup dialogue has value. No matter how much we have a tendency to see groups as unified – the African-American community, the Jewish community, the Serb community – each of these groups consists of individuals and a plethora of views and understandings. Often people cannot effectively come together for dialogue, at least as members of groups, until those groups have had a chance to come together to consider how they wish to interact with the other side. One of the authors was told this in no uncertain terms in Northern Kosovo by Serbs there. Serbs were not ready for interethnic dialogue with Albanians because they hardly knew who they were as a group and up until that point only the loudest, and therefore sometimes the most radical, had had a chance to be part of the discussion. Meanwhile, conflict resolution intervenors were only inviting members of the Serb community who agreed with the international agenda to seminars, in order to avoid "spoilers" and so-called difficult people.

32. *See* CLEVEN, *supra* note 17.

33. The definition in the text is the authors' own usage based on transformative conflict theory. *See* BUSH & FOLGER, *supra* note 16. Broader definitions are found, especially in the literature on social psychology, *see, e.g.,* C. DANIEL BATSON & ADAM A. POWELL, *Altruism and Prosocial Behavior,* HANDBOOK OF PSYCHOLOGY 463 (2003). However, the definition here is generally consistent with those broader ones. The point of the text is not that a facilitator *directs* the parties toward pro-social interaction, but that such interaction is the *result* of the process.

34. *See* AMIN MAALOUF, IN THE NAME OF IDENTITY: VIOLENCE AND THE NEED TO BELONG 12-15 (2003); CHARLES TILLY, IDENTITIES, BOUNDARIES AND SOCIAL TIES 8-9 (2005).

that are shared with other members of their groups. Furthermore, individuals' understanding of their own and others' identities changes as they interact.[35]

When people experience conflict, the transformative model suggests, they become weak and self-absorbed.[36] One important element of the weakness that occurs in conflict is the narrowing of identities: one particular part of who a person is may become most important, even though that person has many other identities. And the part of identity that is shared with other group members becomes all important, at the expense of the others.[37] Self-absorption at the group level is basically polarization, as was already explored in the discussion of political polarization above. People also feel that their group has been uniquely victimized. Sometimes this is of course the case, but often this ignores the reality that victims and aggressors exist on both sides of conflicts.

Dialogue facilitators in other approaches strive for "balance" – working to get similar numbers of people from pre-determined "sides" or groups to participate.[38] Facilitators of transformative dialogue also consider individuals' group membership. But rather than defining who the groups and individuals are in advance, a transformative facilitator will allow those categories and identities to emerge in party-driven fashion, through conversations with members of the community.[39] Our current transformative dialogue work in two communities in the United States provides good examples of the different ways people choose to identify themselves. One community is urban and the other rural.[40] In both, as discussed below, the first steps toward dialogue were initiated by community members themselves, who had heard about transformative dialogue practice, and reached out to transformative facilitators for support.

In the urban community, people themselves define communities in the city in terms of race and ethnicity, and in part according to religion. African-Americans, Hispanics, and various religious denominations define some of the most important groups that people said needed to be involved. This may be because in the urban environment, policing and the interaction between authorities and citizens, to give one example, plays out differently for racial minorities than others. In the rural setting we are working in, political identity is foremost. People contacted us primarily because of a concern about the effects of political polarization on their community. "Trump supporters," "conservatives," and "liberals" are the terms most

---

35. JAMES A. HOLSTEIN & JABER F. GUBRIUM, THE SELF WE LIVE BY: NARRATIVE IDENTITY IN A POSTMODERN WORLD (2000).

36. That is, they lose the clarity they need to make decisions about what they want to do and how they want to relate to others, and they lose the capacity to connect constructively with others and see their perspective. *See* BUSH & FOLGER, *supra* note 16, at 49-53.

37. See MAALOUF, *supra* note 34.

38*. See, e.g.,* Margaret Herzig, *Moving from polarized polemic to constructive conversation–A report from the Public Conversations Project,* 7 J. PUB. PARTICIPATION 1 (2001) (describing the "structured dialogue" process used by one prominent organization, including preplanning of invitations and other "structured" elements described in the text below, *see infra* notes 39-42 and accompanying text); SCHIRCH & CAMPT, *supra* note 28.

39*. See* CLEVEN, *supra* note 17, at 13.

40. The dialogues in these two communities are current and ongoing, and the participants prefer that their work not be cited for publication by name or location, for privacy reasons. Other recent examples of transformative dialogue facilitation include a multi-session dialogue on "passive racism" between white and black residents in a southern Maryland community. *See* Richard "Dusty" Rhoades, *Journey into Self and Other*, *in* TRANSFORMING CONFLICT FROM THE INSIDE OUT 23 (Robert A. Baruch Bush & Joseph P. Folger eds., 2016).

often used to describe different people in this community in spite of the fact that, like their urban counterparts, they also differ in terms of ethnicity and religion. In both urban and rural engagements, the participants self-identified the groups that needed to be involved.

Another practice of transformative dialogue is that the facilitator does not push a particular content or goal for the dialogue. The facilitator does not come in with an agenda to engage participants about a pre-determined topic, but trusts that people know what they want to talk about and what they want to achieve; they also know when they want to change the topic of discussion for some reason. In the rural community we are working in, the concern with political polarization already determines that some topics are of more interest to people than others. In the urban community where we are working, participants requested a meeting several days after the August 2017, Charlottesville incident and its aftermath.[41] Though racism and white nationalism had not been the main focus of earlier sessions, the participants decided to change the topic, the facilitators responded quickly, and people were able to talk about the event's impact on them as individuals and on their community. By contrast, most dialogue processes involve the facilitator in shaping the agenda, either in advance or as the process unfolds, or both. This is another example of the way in which transformative dialogue is genuinely party-driven.

Finally, while most dialogue models give the facilitator the responsibility to set and enforce rules that are supposed to ensure civil exchange, transformative dialogue leaves responsibility for this to the participants themselves – so that heated exchanges and conflict within the dialogue are possible. From such unconstrained exchanges, participants can make lasting changes in their views of themselves and each other.[42] They are free to express their own "truth" and hear others doing the same. In a facilitated face to face exchange individuals have a chance to hear themselves speaking out loud. This often leads people to rethink what they are saying and how they are saying it, and when people do this on their own it leads to real change. Dialogues that have strict rules about how people interact may promote civility while the facilitator is enforcing the rules, but may have little or no long-term effect on interactions. In summary, allowing conversations to go to the heart of the participants' differences allows them to confront what is difficult, to take responsibility themselves for the exchange, and to hear each other's voices in a new way. All of these factors help contribute to real transformation, not simply the polite suppression of difference controlled by a third party.

D. How Transformative Dialogue Can Address the Challenges of Political Polarization

As mentioned earlier, dialogue in a democracy is not a substitute for the political process, nor should it be. In a democratic country, the political process can be

---

41. *See* Aaron C. Davis, Joe Heim and Laura Vozella, *How Charlottesville Lost Control,* WASHINGTON POST (Aug.26, 2017), https://www.washingtonpost.com/investigations/how-charlottesville-lost-control-amid-deadly-protest/2017/08/26/288ffd4a-88f7-11e7-a94f-3139abce39f5_story.html?utm_term=.e932283554d6.

42. *See* Judith A. Saul & Scott Sears, *A Relational Perspective on Multi-Party Practice*, *in* TRANSFORMATIVE MEDIATION: A SOURCEBOOK 397, 407 (Joseph Folger, Robert Bush & Dorothy Della Noce, eds., 2010) (arguing that "Conflict is a prevalent and potentially constructive element of group interaction.").

seen as a form of conflict resolution.[43] Many different opinions and preferences exist with regard to public policy questions, and a country's constitution sets out rules by which people deliberate, debate, advocate and lobby, and finally vote for a particular political representative or policy.

Dispute resolution processes, in order to effectively address the challenges of political polarization, must not supplant this fundamental process. However, they can support and strengthen the political process so that it becomes more constructive in the face of divisions and conflict, even when these conflicts seem intractable.[44] This is true whether or not people reach agreement or reconcile. Transformative processes are uniquely situated to contribute in this way because they do not push people to seek common ground. When conflicts are complex and closely related to identity, a more appropriate goal is to support people in finding ways to interact and relate in spite of their deep disagreements.

There is therefore value and impact in dialogue even when common ground is not sought – in fact we might even say *especially* where common ground is not sought. Transformative facilitators do not define people or their identities in advance. Instead, they follow Levinas in respecting people's radical difference.[45] They allow people to explore their identities, individually or as members or groups. The first main impact of transformative dialogue is that it allows people to choose the identity they wish to emphasize. However, this happens not in an isolated room, but in the real context of that person's community. Through dialogue, people gain clarity and strength to decide how they want to understand and live out their identity. And by interacting with others face to face, the complexity and multi-faceted nature of identity is shaped and constructed more fully.

For example, in one of the community dialogues mentioned above a participant remarked that after talking to a member of the sheriff's department directly, she changed her view of law enforcement officers. As the participant put it, "getting to know someone can change assumptions." Another person remarked that through the dialogue they realized they were more judgmental and angry than they thought they were. These examples attest to the greater clarity in understanding of self and other that emerges in transformative dialogue.

The second potential impact of transformative dialogue is that based on the strength and clarity gained through the process, people can make better decisions going forward as individuals and/or as members of groups. Hearing from others in their own group or from those in other groups often reveals that seemingly simple situations are more complex. And this complexity often opens possibilities that had seemed closed earlier. Sometimes participants may decide not to move forward with certain options because they realize the time is not right. This too is a positive result of a dialogue. Moreover, as mentioned above, a transformative dialogue usually starts through having conversations with individuals as well as small groups to determine who needs to talk to whom, about what, and how.[46] This naturally leads to a focus on local networks. Working with and through local networks rather than bringing individuals together in a pre-determined or random fashion increases the

---

43. *See* Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 363-65 (1978).

44. *See* Heidi Burgess & Guy Burgess, *Constructive Confrontation: A Transformative Approach to Intractable Conflicts*, 13 MEDIATION Q. 305 (1996).

45. *See* LEVINAS, *supra* note 29 and accompanying text.

46. *See supra* note 24 and accompanying text; CLEVEN, *supra* note 17.

possibility of post-dialogue impact, since people are already connected in ways that will continue.

The third potential of transformative dialogue is that it gives people the freedom to recognize the humanity of others.[47] By meeting the other face to face and seeing that they too are not just members of the other group, but unique individuals, and hearing their unique voice and perhaps the voices of other members of their own group in the room, it may be easier to recognize the other's humanity, all the while also recognizing that this does not require agreement or sameness.[48] A dialogue participant can choose to remain "civilly distant" from another.[49] That is still a large step forward from antipathy and hate.[50]

IV. CONCLUSION: TRANSFORMATIVE DIALOGUE AND POLITICAL POLARIZATION

Based on the above discussion, we argue that transformative dialogue can address the challenges of political polarization. First, it does not require common ground. Instead it recognizes that differences exist and that it is all right to "live with no." This makes dialogue more attractive to many because they do not feel threatened by it but see dialogue as a way to establish more constructive interactions across boundaries of definite difference. However, at the same time, the increased clarity and willingness to consider the perspective of others that are hallmarks of this process may sometimes lead to participants finding common ground.

Second, transformative dialogue can alleviate stress and fear because it contributes to recognition of others and their perspectives. As people gain more clarity through the dialogue process, and get to know the other, they also develop strength and confidence. Processes that drive participants towards reconciliation or that emphasize commonalities do not achieve this as effectively, because they are forced, and they often require participants to suppress the way they really want to talk about divisions. Instead, transformative processes confront what is difficult head on and help people deal with it.

Third, transformative dialogue supports people both in finding their voice and what they want to say, and also in being clear about what they do not want to say. The latter is very different from the kind of self-censorship described in the literature on political polarization.[51] If people are self-censoring themselves then they are holding back when they really have things they want to say. Choosing not to speak, from a position of clarity or strength, is different. Still, it may be that in a

---

47. We recognize that other approaches are also aimed at participants recognizing the humanity of others. Our point is simply that controlling the speech that is allowed in a dialogue runs counter to that goal.

48. In fact, as Levinas points out, acknowledgement of this otherness is also an acknowledgement of the freedom we have as human beings.

49. *See* CLEVEN, *supra* note 17, at 10-11 (citing this example: "When interethnic riots broke out, a man [who] was a local nationalist leader among his ethnic group … went out in front of a group of his own people and persuaded them not to commit acts of violence against the other group. He later stated that not only would he not have done that had he not participated in the dialogue process, he would also not even have thought of it. He was not an advocate for reconciliation, but his relationship to members of his own group and members of the other ethnic group had changed [even though] he still had conflict with members of the other group...").

50. *See* AARON T. BECK, PRISONERS OF HATE: THE COGNITIVE BASIS OF ANGER, HOSTILITY AND VIOLENCE (2000) (describing the alienation typical of destructive conflict as a prison).

51. *See supra* text accompanying notes 10-12.

face to face dialogue people find the words they *do* want to say, because as the process progresses it moves from negative and destructive to positive and constructive.

For all of these reasons, discourse has the potential to become more civil in a transformative dialogue and to remain more civil outside the context of the dialogue. Because it allows people to move at their own pace, to speak honestly and in their own words, the changes that occur during a transformative dialogue are more likely to continue when people interact in the future.[52]

Political polarization is a reality in the United States and elsewhere in the world; it has serious consequences that can threaten the way democracy works and therefore also its very legitimacy. The transformative dialogue process can address these challenges. It can help people deal with issues that matter to them and support more constructive conflict interactions, regardless of whether or not people find common ground. In this way, it can create the basis for democratic deliberation within and across group boundaries that can lead to greater tolerance and a better functioning political process – so that even when we cannot get to yes, we can nevertheless live civilly with no.

---

52. *See supra* text accompanying notes 45-46.

5-1-2017

A Divided Nation: Political Polarization and Dispute Resolution

Lindsey Phipps

A Divided Nation: Political Polarization and Dispute Resolution

Lindsey Phipps

Abraham Lincoln stated in 1838, "America will never be destroyed from the outside. If we falter and lose our freedoms, it will be because we destroyed ourselves."[1] This statement seems to have predicted the demise we are faced with today, which is, American political polarization. Our nation is so divided, that even in the face of a crisis our legislative body cannot compromise.[2] Altogether, this has halted the productivity of the legislative process and diminished the effective governance of our democracy.[3] As political figures increasingly trade up patriotism for partisanship, one may question whether productivity will cease to exist in a politically divided government body.[4]

Moreover, if voters are hoping for a politically homogenous legislative, judicial, and executive branch, for the sake of productivity, what does that mean for democracy? In the words of Bill Moyers, a former White House Press Secretary, "[a]lthough our interests as citizens vary, each one is an artery to the heart that pumps life through the body politic, and each is

---

1. *Lincoln's Legacy: A Conversation with Henry Louis Gates, Jr.*, PBS (Jan. 12, 2009), http://www.pbs.org/wnet/lookingforlincoln/about/lincolns-legacy-aconversation-with-henry-louis-gates-jr/297/.

2. The 2013 government shutdown exemplifies the inability and unwillingness of legislative officials to make compromises in an emergency. In 2013, U.S. national debt increased to $17 trillion so Congress enacted the sequestration, which included a number of spending cuts. With the new fiscal year and the implementation of the Affordable Care Act approaching, Congress was pressured into creating a resolution. However, the Democrat-led Senate and the Republican led-House could not agree on a resolution leaving the government officially shut down by October 1st. Brittany Crocker, *The Government ShutDown Explained*, Global Journalist (Oct. 24, 2013), http://globaljournalist.org/2013/10/the-government-shutdown-explained/.

3. *Id.*

4. *See* James R. Rogers, The Impact of Divided Government on Legislative Production, 123 Public Choice 217, 225-26 (2005).

> [D]ivided legislatures have a large and statistically significant impact on legislative production. Each chamber in a divided legislature on average originates 80 fewer laws relative to chamber in unified legislatures. That represents a decrease in legislative production of almost 30 percent relative to the average number of enactments by chambers in unified legislatures.

Id.

111

important to the health of democracy."[5] The Framers intended for both majority and minority opinions to be safeguarded by our system of checks and balances—safeguards that come at a high cost today.[6]

Because we cannot conceive of relinquishing our democratic principles of checks and balances in exchange for a productive government body, we must pressure elected officials to find common ground with their adversaries. This requires that elected officials shift their attitudes from one of competitive to cooperative. They must expect mutually satisfactory outcomes rather than a zero-sum result. Though our country thrives on a politically adversarial discourse, our elected officials are ultimately playing on the same team. As President Barack Obama stated, "[t]he strongest democracies flourish from frequent and lively debate, but they endure when people of every background and belief find a way to set aside smaller differences in service of a greater purpose."[7] America's democratic principles and need to carry out efficient governance greatly outweighs any small victories achieved by stubborn political parties.

This article will discuss the causes and consequences of party polarization and propose that the legislative body work more collaboratively and cooperatively through direct implementation of alternative dispute resolution techniques such as negotiation and mediation in the legislative process. Part I will define political polarization, what it looks like today, its causes and its consequences. Part II will propose and explain the use of dispute resolution techniques and tactics, such as, negotiation, mediation and alternative dispute resolution to mitigate the effects of political polarization. Part III will conclude that dispute resolution techniques and tactics will mitigate the inadequacies created by political polarization.

PART I: WHAT IS POLITICAL POLARIZATION?

Political polarization refers to "the vast and growing gap between liberals and conservatives . . . [which is] a defining feature of American politics today."[8] It is the "excessive partisanship and deep ideological

---

5. BILL D. MOYERS, MOYERS ON DEMOCRACY 14 (2008)

6. Paul M. Johnson, *Checks and Balances*, A GLOSSARY OF POLITICAL TERMS, https://www.auburn.edu/~johnspm/gloss/checks_and_balances (last visited Sept. 20, 2016).

7. Macon Phillips, *The First Press Conference*, THE WHITE HOUSE (Feb. 9, 2009, 8:10 PM), https://www.whitehouse.gov/blog/2009/02/09/first-press-conference.

8. Political Polarization, Pew Research Center, http://www.pewresearch.org/packages/political-polarization/ (last visited Sept. 20, 2016).

112

division among political elites and officeholders."[9] Political polarization has defined the political climate since the 1970's, and the divide continues to increase.[10]

Surprisingly, political parties were never intended by our founding fathers; they developed merely by accident.[11] In fact, the Framers warned of their destructive nature.[12]   For instance, President George Washington stated:

> However [political parties] may now and then answer popular ends, they are likely in the course of time and things, to become potent engines, by which cunning, ambitious, and unprincipled men will be enabled to subvert the power of the people and to usurp for themselves the reins of government, destroying afterwards the very engines which have lifted them to unjust dominion.[13]

Despite these warnings, political parties formed almost immediately in the late 1790s.[14] Although undermining the central premise of democracy ("the people will rule"), the creation of political parties bridged that gap between elected officials and their voters as "vehicles of promotion."[15]

Beginning in the early 1960's, Franklin Roosevelt's Old New Deal Coalition ("The Coalition") began to dissipate.[16]   The Coalition had maintained the dominance of the Democratic Party in years prior.[17]   It consisted of a diverse group of voters with little in common, ranging from white southerners to ethnic minorities and working class voters.[18] With the

---

9.   Michael Barber & Nolan McCarty, *Causes and Consequences of Polarization*, *in* Negotiating Political Agreements, AM. POL. SCI. ASS'N 19, 19 (Jane Mansbridge & Cathie J. Martin eds., (2016) http://www.apsanet.org/portals/54/Files/Task%20Force%20Reports/Chapter2Mansbridge.pdf.

10. Drew DeSilver, *The Polarized Congress of Today has its Roots in the 1970s*, PEW RESEARCH CENTER (June 12, 2014), http://www.pewresearch.org/fact-tank/2014/06/12/polarized-politics-in-congress-began-in-the-1970s-and-has-been-getting-worse-ever-since/.

11. Todd Philips, *Political Parties Were Never Meant To Be,* THE HUFFINGTON POST (Aug. 31, 2012, 1:32 PM), http://www.huffingtonpost.com/todd-phillips/political-parties-were-ne_b_1846903.html.

12. *Id.*

13. *George   Washington's   Farewell   Address   1796*, YALE   LAW   SCHOOL, http://avalon.law.yale.edu/18th_century/washing.asp (last visited Oct. 13, 2016).

14. *Two Parties Emerge*, U.S. HISTORY, http://www.ushistory.org/us/19c.asp (last visited Sept. 20, 2016).

15. Philips, *supra* note 11.

16. Alan Abramowitz, *How did American Politics Become So Polarized*, BIG THINK, http://bigthink.com/videos/how-did-american-politics-become-so-polarized (last visited Sept. 20, 2016).

17. *Id.*

18. *Id.*; *see also Franklin D. Roosevelt: The American Franchise*, MILLER CENTER, http://millercenter.org/president/biography/fdroosevelt-the-american-franchise (last visited Sep. 20,

113

increase of African American voters after the enactment of the Voting Rights Act, the Coalition became unstable as its more conservative members moved toward the Republican Party,[19] forging an ideological realignment that influenced the next several decades.[20]

The last several decades of voting records (1970-2012) evidence this ideological shift,[21] from substantial overlap of the parties in the House of Representatives,[22] to essentially no overlap in either the House or the Senate.[23] Currently, political antagonism has taken control of Congress, causing the legislative process to resemble a never-ending game of tug-of-war.[24] For instance, "voting in Congress is now almost purely one-dimensional— [political ideology] accounts for about ninety-three percent of roll call voting choices in the 113th House and Senate."[25] This trend is also mirrored by the public's attitude towards political rivals,[26] as evidenced by one-sided voting patterns, which have doubled over the last decade.[27] Consequently, any ideological commonality between the two parties has decreased substantially.[28] In addition, the moderate platform has decreased as voters who identify as moderates still maintain very conservative or

---

2015). "With FDR at its head, the Democratic Party put together a formidable coalition whose main components were lower-income groups in the great cities—African-Americans, union members, and ethnic and religious minorities, many from recent immigrant groups—and the traditional source of Democratic strength, 'the Solid South.'" *Id.*

19.   Abramowitz, *supra* note 16.

20.   *Id.* The ideological realignment became apparent in the Nixon campaign, and continued into the Reagan campaign and into the take over of the House by the Republican Party led by Newt Gingrich. *Id.*

21.   DeSilver, *supra* note 10.

22.   *Id.* In 1973-74, 240 House members scored in between the most conservative Democrat congressman and the most liberal Republican congressman. In the Senate, twenty-nine senators scored between the most liberal Republican senator and the most conservative Democrat senator. By 1983-84, those numbers decreased to only ten senators and sixty-six representatives. Again, by 1993-94, that number fell to nine congressmen and three senators. *Id.*

23.   *Id.*

24.   *Id.*

25.   *Id.*

26.   Carroll Doherty, *7 Things to Know about polarization in America,* PEW RESEARCH CENTER (June 12, 2014), http://www.pewresearch.org/fact-tank/2014/06/12/7-things-to-know-about-polarization-in-america/.

27.   *Id.* "The share of Americans who express consistently conservative or consistently liberal opinions has doubled over the past two decades, from 10% to 21%." *Id.*

28.   *Id.* "The 'median,' or typical, Republican is now more conservative than 94% of Democrats, compared with 70% twenty years ago. And the median Democrat is more liberal than 92% of Republicans, up from 64%. Among Republicans and Democrats who are highly engaged in politics, 70% now take positions that are mostly or consistently in line with ideological bent of their party."

114

liberal opinions on hot-button issues such as gun control, health care policy, and immigration.[29]   Moreover, this political divide among the general population has gone beyond political ideology and has developed into political antipathy.[30]   Even so, members of the public view their opposing party as more than just an adversary; rather, they are viewed as a threat to national security.[31]

The negative consequences of political polarization manifests itself in Congress' inability to pass substantive laws.[32]   By the end of Congress' term in 2014, the 113th Congress had passed 212 pieces of substantive legislation, compared to the 463 pieces of substantive legislation passed by the 106th Congress, just a decade before.[33]   Polarization threatens to destroy the efficiency of our democratic system[34] by complicating our government's ability to address domestic policy problems, halting the implementation of foreign policy and national security strategy, eroding public trust in the government, and causing intolerable, long-lasting destruction to American institutions.[35]

Traditionally, compromise has been the solution to getting legislation passed, however, it has become increasingly difficult due to the current political climate.[36]   For one, polarized representatives are inherently reluctant to compromise.[37]   Further, constituents are unlikely to pressure their elected officials to compromise in an effort to get legislation passed.[38] Although voters say they prefer political officials who are willing to cooperate, they still prefer an end result favorable to their interests.[39]

---

29.   *Id.*

30*.   Id.* "The share of Republicans who have very unfavorable opinions of the Democratic Party has jumped from 17% to 43% in the last 20 years. Similarly, the share of Democrats with very negative opinions of the Republican Party also has more than doubled, from 16% to 38%." *Id.*

31.   *Id.*

32.   Drew DeSilver, *In Late Spurt of Activity, Congress Avoids 'Least Productive' Title*, PEW RESEARCH Center (Dec. 29, 2014), http://www.pewresearch.org/fact-tank/2014/12/29/in-late-spurt-of-activity-congress-avoids-least-productive-title/.

33.   *Id.*

34.   William A. Galston, *Political Polarization and the U.S. Judiciary,* 77 UMKC L. REV. 307, 319-20 (2008).

35.   *Id.* at 319-20.

36.   Drew DeSilver, *The Polarized Congress of Today has its Roots in the 1970s*, PEW RESEARCH CENTER (June 12, 2014), http://www.pewresearch.org/fact-tank/2014/06/12/polarized-politics-in-congress-began-in-the-1970s-and-has-been-getting-worse-ever-since/.

37*.   Id.*

38*.   Id.*

39*.   Id.* "A 56% majority prefers political leaders who 'are willing to compromise,' while 39% prefer leaders who 'stick to their positions.'" *Section 4: Political Compromise and Divisive Policy*

115

There are several different theories as to why American politics have become so polarized. These theories can be divided into two categories: external and internal.[40] The external category emphasizes movement in the societal, economic, and electoral environments, and how they have altered incentives for political officials to cooperate or polarize.[41] The internal category highlights the formal and informal institutions of Congress and how they have transformed it in ways that increase polarization.[42] It is important to highlight these theories because they reveal inherent problems within the legislative system that can be addressed specifically by dispute resolution tactics. Additionally, sorting out the reliable and unreliable theories will help in identifying dispute resolution tactics as a remedy that is narrowly tailored to this specific problem.

External Category

Under the external category is the "Polarized Electorate" theory, which states that elected officials are polarized because their constituents are.[43] This theory, though enticing, has little evidence to support it.[44] It requires the evidence of two trends: (1) *partisan sorting*, which is the increasing support for political parties based on ideology, and (2) an increasingly polarized policy preference or ideological identification.[45] As to partisan sorting, there is evidence that voters have become better sorted into the party system.[46] However, as to the second trend, voter-policy polarization is less clear.[47] Though voters identify with a party, their policy positions remain moderate.[48] In fact, representatives were more extreme in their policy positions than their constituents.[49] Although this theory lacks substantial evidence to support it, it demonstrates that voters will sort themselves according to elite polarization so the incentives for representatives to take

---

Debates, PEW RESEARCH CENTER, (June 12, 2014), http://www.people-press.org/2014/06/12/section-4-political-compromise-and-divisive-policy-debates/.
40.   Barber and McCarty, supra note 9, at 23.
41.   *Id*.
42.   *Id*.
43.   *Id*.
44.   *Id*.
45.   *Id*. at 23-24.
46.   *Id*. at 24. "[O]ver time, voters have increasingly held political views that consistently align with the parties' policy positions." *Id.*
47.   *Id*. at 25.
48.   *Id*. at 25.
49.   *Id*.

116

positions that support their constituents will reduce.[50] Thus, public pressure alone may have little effect in implementing any mediation or negotiation strategies in legislative making.

Another theory under the external category is the "Southern Realignment," which states that because conservatives now dominate the historically liberal South, median southern Democrats are now more liberal and the southern Republicans have become more conservative.[51] However, this theory lacks support because it relies primarily on the disappearance of conservative Democrats.[52] Doing so neglects the polarization of non-Southerners.[53] Still, this movement shows us that the change in demographics of the Southern electorate has impacted polarization in the South.[54]

Moreover, "gerrymandering" has also been cited as an external explanation for the polarization in American politics.[55] Gerrymandering is defined as "the dividing of a state, county, etc., into election districts so as to give one political party a majority in many districts while concentrating the voting strength of the other party into as few districts as possible."[56] The argument is that because state legislatures are permitted to draw congressional district lines, districts are becoming overwhelmingly partisan and prevent other parties from competing against the dominant party.[57] Yet the evidence supporting this theory is weak.[58] First, this does not apply to the Senate since they are elected by the entirety of the state.[59] However, evidence has shown that Congress would be just as polarized if districts were drawn randomly.[60] Thus, the argument that the difference in how Republicans and Democrats represent moderate districts relates more to polarization than the difference within a party representing a party-

---

50. *Id.*

51. *Id.* at 26*; see also* Abramowitz, *supra* note 16.

52. Barber & McCarty, *supra* note 9, at 26.

53. *Id.*

54. *Id.* at 27.

55. *Id.*

56. *Gerrymander,* DICTRIONARY.COM   http://dictionary.reference.com/browse/gerrymander (last visited Sep. 21, 2016, 10:33 AM).

57. Barber & McCarty, *supra* note 9, at 27.

58. *Id.*

59. *Id.*

60. Barber & McCarty, *supra* note 9. "McCarty, Poole and Rosenthal (2009) generated random districts and determined the expected partisan ship of representatives form these hypothetical districts given the demographic characteristics of the simulated district. The result was that the simulated legislatures generated by randomly creating districts are almost as polarized as the current Congress." *Id.*

117

dominated district.[61] Again, this moves the emphasis away from the public to the legislative representatives when implementing mediation and negotiation tactics.

Further, many scholars rely on primary elections to explain the polarization in American politics.[62] In a primary election, registered party members vote on a candidate to represent their party in the general election.[63] Scholars argue that moderate candidates should be allowed to participate in partisan primary elections since "only conservatives can win Republican primaries and [only] liberals can win Democratic primaries."[64] Though there has been some supportive evidence to back this argument, it is moderate at best.[65] Still, the recent change in California from a closed primary to a blanket primary, in which all voters can participate, shows some hope in producing moderate candidates in competitive districts.[66] Further, this highlights the affects the implementation of a process designed to encourage bipartisanship and provides support for implementing an alternative dispute resolution procedure in the legislative process.

The argument that economic inequality has resulted in polarization is more promising.[67] The relationship between polarization and economic inequality is unique.[68] Inequality caused by rising high incomes supports the increase in conservative economic policies, which then discourages policy responses to economic inequality, which in turn facilitates greater inequality causing polarization.[69] The growing divide between the rich and the poor,

---

61.   Barber & McCarty, *supra* note 9, at 27.

62.   *Id.*

63.   *How Does the Presidential Primary Process Work?*, VOTESMART, https://votesmart.org/education/presidential-primary#.VlNF6cbA20s (last visited Sep. 21, 2016).

64.   Barber & McCarty, *supra* note 9, at 29.

65.   *Id.*

Kaufmann, Gimpel, and Hoffman (2003) found that presidential primary voters in states with open primaries hold political ideologies similar to the general electorate, whereas in states with closed primaries, the two electorates are more ideologically distinct. Gerber and Morton (1998) found that the positions of legislators nominated in open primaries hew more closely to district preferences, whereas Brady, Han, and Pope (2007) found that legislators who hew closely to the general-election electorate suffer an electoral penalty in primaries.

Id.

66.   Barber & McCarty, *supra* note 9, at 29.

67.   *Id.* at 30.

68.   *Id.*

69.   *Id.* "[V]oting behavior and partisan identification increasingly correlate with income . . . and that the ideal points of legislators are increasingly correlated with average district income."

118

and its relationship to polarization, provide a valuable insight into its effect on legislation produced by the legislative body.[70]

Another theory suggests that private campaign finance is linked to polarization.[71] The argument is that elected officials take extreme policy positions on behalf of their donors.[72] Still, the evidence supporting this theory is weak.[73] Though research suggests that individual donors are more extreme than non-donors, and the amount of individual donors is increasing, there is still insufficient evidence to support the connection.[74] Even so, it does not readily imply a causal connection.[75]

The media's effect on politics is a familiar topic, and thus, a common theory for why American politics has become so polarized.[76] It is not news that media outlets pick sides. So too, do their viewers.[77] For this reason, elected officials and candidates fear that if they take moderate stances, party activists will use that against them.[78] Additionally, these media outlets give voice to extreme political activists.[79] This theory is significant considering that America relies on transparency to facilitate democratic self-governance and conversely, mediation relies on private conversation.[80]

Internal Theories

Under the internal explanations category is the "rule change" theory, which states that polarization is due to changes in the rules and procedures of Congress.[81] More specifically, the amendment process was modified to encourage party line voting.[82] Yet, this argument is deficient for several

---

70. *Id*. at 30.

71. *Id*. at 31-32.

72. *Id*. at 32.

73. *Id*.

74. *Id*. ; *see also* Doherty, *supra* note 25. "Political donation rates are roughly double the national average among ideologically consistent liberals (31% have donated money) and conservatives (26%)." *Id.*

75. Barber & McCarty, *supra* note 9, at 32.

76. *Id*.

77. *Id*.

78. *Id*.

79. *Id*.

80. Barack Obama, *Transparency and Open Government*, THE WHITE HOUSE, https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment (last visited Sept. 21, 2016); Adam L. Powers, *Frequently Asked Questions About Mediation*, MARK POWERS LLP, http://www.markspowers.com/pg15.cfm (last visited Sept. 21, 2016).

81. Barber & McCarty, *supra* note 9, at 33.

82. *Id*. at 34.

119

reasons: (1) change does not affect the Senate, (2) it is unlikely that a single rule change can explain a gradual increase in polarization over the last four decades, and (3) it is not distinct from the procedures in state legislatures.[83] Another similar argument states that the power of the controlling party in the House has increased their ability to generate more party line votes.[84] However, like the rule change theory, this only applies to the House and does not account for the strength of the growing polarization.[85] Still, this theory demonstrates the effects of procedure on deal-making, which would support implementing a dispute resolution process to legislative procedure.

Further, scholars argue that because party leaders in the House and Senate have become so powerful, they can pressure members of the party to vote on party lines, thereby increasing polarization.[86] This argument fails because party pressure cannot be measured without making strong assumptions about the behavior of individual elected official.[87] Similarly, the "teamsmanship" theory is commonly used to explain the polarized political climate.[88] The argument is that the competitive nature of the elections drives party members to increasingly distinguish themselves from their opposite party creating a teamsmanship-like environment.[89] Still, this theory does not account for the sudden up rise in polarization in our country.[90] Under this theory, polarization should have begun and stuck around as soon as political parties gained traction in the electorate.[91] Even so, this argument highlights the complexity of law making and the need to implement a mediation policy that would represent the interests of all party members, not just the present parties to the negotiation.

---

83. *Id*. at 33-34.
84. *Id*. at 34.
85. *Id*.
86. *Id*.
87. *Id*.
   "Snyder and Groseclose assumed that members are free from party pressure on lopsided votes; therefore, a comparison between positions on lopsided and close votes can reveal the effects of party pressure. They found that, indeed, there are policy areas in which party pressure is more common, but they did not find a steady increase in partisan pressure commensurate with the increase in polarization observed during the past 40 years."
Id.
88. *Id*. at 35.
89. *Id*.
90. *Id*.
91. *Id*.

120

Lastly, it is believed that the breakdown of bipartisan norms in Capitol Hill has led to polarization.[92] As the ability to travel has increased, elected officials no longer need to stay in Washington D.C. when they do not have to; instead, they can travel back to or stay in their home states.[93] Because of this, representatives are less capable of creating cross-partisan friendships with their co-members causing a lack of the trust and civility necessary to make compromises.[94] Though this hypothesis is tempting, it has not been substantiated.[95] Yet this theory signifies the legislature's need for trust and camaraderie and the lack there of. Mediation and negotiation can play a significant role in abating this problem by highlighting the need for mutual understanding and respect between the parties.[96]

So, which theory is correct? In answering this question, it is important to remember that these theories are not mutually exclusive; external and internal forces can combine to cause and exacerbate the problem.[97] Thus, the focus should not be on identifying a sole cause of polarization, but rather on what these theories together say about the problem and the solution. Seeing and understanding the full picture of the problem makes it easier in identifying a practical, workable solution.

PART II: WHAT IS THE SOLUTION?

Turning to the solution of political polarization, how do we begin to solve this crisis? "As members of the ADR profession, we have a responsibility to bring our skills to the forefront to help develop civil collaborative discussion (a) among us as citizens, (b) between citizens and politicians, and (c) push for it among our politicians."[98] Following this call

---

92. *Id.*

93. *Id.*

94. *Id.*

95. *Id.*

96. S. Elise Peeples, *The Space Between the Parties: A Model for Understanding the Magic of Mediation*, MEDIATE (Dec. 1998), http://www.mediate.com/articles/peeples1.cfm.

> We do not come together to determine right or wrong, winners or losers, but to attempt to express, to receive and to respect. If respect and understanding occur, then it is not essential that the two parties agree with each other—it is enough that they are able to move into the Between . . . where we can let go of ego . . . and find the solutions that have evaded us.

Id.

97. Barber & McCarty, *supra* note 9, at 23.

98. Craig Tipton Distelhorst, *Civil, Collaborative Discourse on Public Policy*, 45 MD. BAR J. 38, 40 (2012).

121

to duty, the question then becomes, how can dispute resolution techniques solve the issue? In the question lies the answer: dispute resolution techniques resolve disputes, such as the ones that exist in the American legislature. "[N]egotiators often fail to reach agreements [sic] because they lack a process for exploring and meeting the essential interests of both (or all) parties."[99] Dispute resolution techniques can provide legislators with that "process," which could abate the polarization crisis.[100]

First, legislators must employ "deliberate negotiation," that is, "negotiation based on a process of mutual justification, respect, and reciprocal fairness."[101] This type of negotiating would require legislators to "(1) [i]dentify win-win issues to make strategic trades, (2) look beyond the 24/7 news cycle, (3) seek out low-risk venues for honest, open communication, (4) sequence issues strategically to build trust, and (5) engage genuine differences with vigor and principle."[102]

The first tactic Congress must employ is seeking mutually satisfying options in bargaining.[103] Ideally, a win-win agreement would be one that "each party gains but with distinct benefits; no losses."[104] Where these agreements can be found, they should be made; however, more commonly, legislators find themselves in a "Negotiation Myopia."[105] A "Negotiation Myopia" is a scenario in which both parties differ too greatly to make a good deal for both of them, but other promises can be added to the deal to satisfy both parties.[106] However, in the legislative process, there are many more interested parties to the deal, such as party members and constituents, both complicating the deal.[107] Still, "mutual gains may be discovered or created through negotiations [and] . . . these gains often build on taking the

---

99. L. Michael Hager, *Congress Needs a Mediation Tool to Dissolve Gridlock* (June 18, 2010), THE WASHINGTON POST, http://www.washingtonpost.com/wp-dyn/content/article/2010/06/17/AR2010061704566.html (alteration in original) (quoting ROGER FISHER, WILLIAM L. URY, & BRUCE PATTON, GETTING TO YES (2011)).

100. *Id.*

101. Mark E. Warren & Jane Mansbridge, *Deliberative Negotiation*, *in* Negotiating Political Agreements, AM. POL. SCI. ASS'N 86, 92 (Jane Mansbridge & Cathie J. Martin eds., 2016) http://www.apsanet.org/portals/54/Files/Task%20Force%20Reports/Chapter5Mansbridge.pdf.

102. Robert C. Bordone & Tobias Berkman, *Negotiation Advice for the 112th Congress*, HARVARD L. BLOG (Nov. 17, 2010), http://www.hnlr.org/2010/11/negotiation-advice-for-the-112th-congress/.

103. *Id.*

104. Warren & Mansbridge, *supra* note 101, at 93.

105. *Id.* at 95.

106. *Id.*

107. *Id.* at 96.

122

perspective of others; and . . . those perspectives often can be obtained in the course of informal, friendly, repeated, and relatively open relationship."[108]

The notion that "[c]ommunication is at the root of any successful negotiation"[109] should be a common, reoccurring theme in negotiation strategy.[110] "[R]epeated interactions in Congress . . . is a 'key ingredient to ongoing successful negotiation.'"[111] Even so, the complex nature of the legislative process adds several other parties to the mix.[112] Constituents and elected members of the political party must feel like their interests are being represented in these deals too.[113] Thus, it is important to not only encourage a strong relationship between both parties present to the deal, but it is also important to encourage a relationship with constituents and other elected members of the respective political party.[114]

This may be difficult because, as mentioned in the third strategy above, parties must "seek out low-risk venues for honest, open communication," which implicates private meetings behind closed doors.[115] "[T]hese settings allow for arguing and persuasion, because negotiators do not have to stick to their fixed preferences behind closed doors and are allowed to 'think out loud' about possible negotiating solutions."[116] However, in this context, "closing the doors" necessarily assumes that one or more parties will not be in attendance, namely, constituents or other elected members of the respective political party.[117] This contradicts our notions of a democratic society, transparency being key to self-governance.[118] Yet, delving more into the topic reveals how transparency is not always "democratic."[119]

The second strategy above—"look beyond the 24/7 news cycle"[120]—recommends that Congressmen look past their representation in the media

---

108. *Id.* (alteration in original); *see also* Bordone & Berkman, *supra* note 102. "There are other sets of issues where one party might care a lot and the other party might be largely indifferent . . . By trading on such issues, both parties gain."

109. Glenn Davis, *Why Can't Congress Negotiate?*, INDEP. VOTER NETWORK (Jan. 27, 2014), http://ivn.us/2014/01/27/cant-congress-negotiate/.

110. Bordone & Berkman, *supra,* note 102.

111. Davis, *supra* note 109.

112. Warren & Mansbridge, *supra* note 101, at 96.

113. *Id.*

114. *Id.*

115. Bordone & Berkman, *supra* note 102.

116. Warren & Mansbridge, *supra* note 101, at 106.

117. *Id.*

118. Warren & Mansbridge, *supra* note 101, at 108.

119. *Id.*

120. Bordone & Berkman, *supra* note 102.

123

because it invites the undue influence of those seeking to have their voice heard.[121] Political activists are those who pay the most attention to those detailed negotiations, and because of this, their voices are heard louder than other constituents.[122] Further, this was not a norm perceived by our founding fathers, as even "delegates at the Federal Convention debated whether to require that roll-call votes be made public."[123]

There is clear evidence from the United States and the United Kingdom that "demands for transparency appeared during periods of heightened fears that representatives were biased."[124] Conversely, "during periods where fears of bias were less present, the public was more accepting of closed-door sessions."[125] Thus, the desired setting for having deliberative negotiations can be created by building trust with the respective elected political party members and constituents.[126] Developing relationships can build trust.[127] Yet, constituents pose a unique challenge to developing relationships. To develop a strong relationship with a constituency, an elected official must "(1) respect and understand the needs and motivations of [his/her] constituents, (2) meet their needs if they are in keeping with your organization's values, mission, and vision, (3) communicate your programs and activates and their value to the constituents, and (4) follow up and nurture the relationship."[128] Thus, constituents might consider looking for these behaviors in re-electing an elected official for another term, and conversely, elected officials might seek to implement these behaviors to win the trust of the public.

Further, it may be necessary to consider the effect incumbency has on relationships in the political arena.[129] Incumbency may allow for more repeated interaction between members of the opposing party that can fuel a better relationship for making deals.[130]

---

121.   Warren & Mansbridge, *supra* note 101, at 109.

122.   *Id*. at 110.

123.   *Id*. at 109.

124.   *Id*.

125.   *Id*.

126.   Warren & Mansbridge, *supra* note 101, at 111.

127.   Yan Ki Bonnie Cheng, *Power and Trust in Negotiation and Decision-Making: A Critical Evaluation*, HARV. NEGOT. L. REV. (Sept. 1, 2009), http://www.hnlr.org/2009/09/power-and-trust-in-negotiation-and-decision-making-a-critical-evaluation/.

128.   Simone P. Joyaux , *Building Relationships with Your Constituents*, SimoneJoyaux 1, 1 http://www.simonejoyaux.com/downloads/BuildingRelationships.pdf (last visited Sep. 22, 2016).

129.   Warren & Mansbridge, *supra* note 101, at 104.

130.   *Id*. at 105.

124

> When adversaries know each other well, they are far more likely to recognize whether the other side's refusal to compromise on a principle is a negotiating tactic or a real political constraint. They are less likely to act as players in the classic bargaining game who hold out for their maximum individual payoff, producing an outcome that makes both sides worse off. In longer-term relationships, legislators have a better sense of their colleagues' intentions, their trustworthiness, and the political constraints they are facing- and their colleagues know that they do. They are repeat players. That enables all to make more confident judgments about when to compromise and when not to.[131]

Still, this contradicts democratic norms because incumbents generally have less-contested elections and the public loses control over their representatives when they do not have a quality competitor, or a competitor at all, to vote for in the alternate.[132] However, in light of the relationship between incumbencies and deal making, the public should keep this in mind when voting for incumbents, as should incumbents when they decide their campaign strategy.

Even where relationships fail or prove to be insufficient, trust can be built. Consider the fourth strategy, which recommends that legislators "sequence issues strategically to build trust."[133] This would require legislators to make small, mutually satisfactory agreements at the outset of the negotiation before moving on to more difficult issues.[134] This would allow legislators to "generate positive momentum and tangible gains" and makes the adversarial party "less likely to defect later when the issues become more challenging."[135] Though this may seem contradictory to the fifth strategy recommendation to "engage genuine differences with vigor and principle," it is important to remember that "[k]eeping these differences from spilling over and poisoning progress in other domains, however, is a hallmark of a truly great negotiator."[136] In other words, legislators would be well advised to "pick and choose their battles" because some small issues are worth conceding on to "lose the battle but win the war."[137]

Mediation can also be an effective and practical tool for our polarized legislatures. In such situations, "[w]hen senators or representatives find themselves locked into irreconcilable positions on issues of national

---

131. Warren & Mansbridge, *supra* note 101, at 104-05.
132. *Id*. at 105.
133. Bordone & Berkman, *supra* note 102.
134. *Id*.
135. *Id*.
136. Bordone & Berkman, *supra* note 102.
137. *Id*.

125

importance, third-party mediation could help overcome a stalemate."[138] A neutral, third-party can do this by looking at the issue in a nonpartisan manner, which will help the parties move away from their party-allegiances and look at the issue objectively.[139] This comes at no risk to legislators because, unlike arbitration or litigation, mediations are simply "assisted negotiations."[140] In other words, "[t]he mediator assists the parties to understand their respective interests and helps them devise solutions that satisfy those interests—but only the parties can decide the ultimate outcome."[141]

Still, this can be easier said than done because elected officials are subject to public scrutiny.[142] For instance, if a legislator takes a hard position on an issue during his campaign, he may feel constrained in altering his position.[143] Still, in that case, mediation will at least provide legislators the opportunity to have their side considered, and to hear and consider the other party's position. That way, the legislator's supporters can feel like their opinion was heard and considered too, while also having the opportunity to reconsider their own position.

Mediation is also especially well-suited for addressing *all* parties involved (party members and constituents) because of the opportunity to "caucus."[144] During a caucus, each interested party can meet separately with the mediator allowing the party to maintain a hard position while exploring other options.[145] This is similar to the "closed door" scenario referred to earlier, where legislators can talk more freely without the public scrutiny.[146]

---

138. Hager, *supra* note 99; *see also* Sarah Gonski, *Easing Gridlock in the United States Congress through Mediation: Letting our Cities and Streets Teach Us Lessons on Getting Along*, THE A.B.A., 1, 2 (2013), http://www.americanbar.org/content/dam/aba/events/dispute_resolution/lawschool/boskey_essay_contest/2013/easing_gridlock_in_the_united_states_congress_through_mediation_letting_our_cities_and_states_teach_us_lessons_on_getting_along.authcheckdam.pdf.

139. *See* Hager, *supra* note 99.

140. *Id.*

141. *Id.*

142. *See* Norm Brand, *Using Mediation Effectively,* MEDIATE.COM, (Sept. 2000), http://www.mediate.com/articles/nbrand2.cfm.

143. *Id.*

144. *Id.*

145. *Id.*

146. Sarah Gonski, *Easing Gridlock in the United States Congress through Mediation: Letting our Cities and Streets Teach Us Lessons on Getting Along*, THE A.B.A., 1, 19 (2013), http://www.americanbar.org/content/dam/aba/events/dispute_resolution/lawschool/boskey_essay_contest/2013/easing_gridlock_in_the_united_states_congress_through_mediation_letting_our_cities_and_states_teach_us_lessons_on_getting_along.authcheckdam.pdf.

126

Complicating the situation is the fact that confidentiality is a large component of mediation and contradicts our notions of democracy, which values transparency.[147] However, in this case, the confidential conversations going on behind closed doors are with a mediator, rather than the opposing party.[148] Because of the mediator's neutral nature and lack of authority to decide an outcome, there are little public concerns that justify transparency in this case.[149] Even so, mediators should be required to have "a politically neutral history; campaign donations or public statements made in favor of one party or another would necessarily disqualify them for employment by inspiring distrust or suspicions of bias."[150]

Overall, the democratic concerns are substantially outweighed by the benefits of mediation. Mediation has the potential to (1) "increase the number of truly bipartisan compromise bills,"[151] (2) "move legislation toward the middle of the political spectrum,"[152] and (3) "[b]etter organiz[e] for complex issues."[153] If anything, mediation should become a "democratic norm" because it reconciles with the vision of the country's founders.[154]

Further, states already use mediators in their legislative process, which could serve as a model for the federal government.[155] For instance, the Illinois General Assembly asked Judge Getty to lead a multi-party mediation in an effort to create a successor to a bill that was within hours of expiring.[156] The expiration of the bill would have meant that the telecommunications industry would go unregulated.[157] Even this looming expiration date wasn't enough to encourage agreement between the opposing political parties.[158] Thus, mediation served as the key to creating a bill that was accepted almost unanimously before time ran out.[159]

---

147. *Id*. at 14.

148. *See* Brand, *supra* note 142.

149. *Id*.

150. Gonski, *supra* note 146, at 12.

151. *Id*. at 16.

152. *Id*. at 17.

153. *Id*. at 19.

154. *See* Todd Philips, *Political Parties Were Never Meant To Be,* THE HUFFINGTON POST (Aug. 31, 2012, 1:32 PM), http://www.huffingtonpost.com/todd-phillips/political-parties-were-ne_b_1846903.html.

155. Gonski, *supra* note 146, at 4.

156. *Id*.

157. *Id*. "Because regulation of the telecommunications industry is such a contentious issue with significant amounts of money at stake, over thirty interested parties were invited to participate in mediation." *Id.* at 5.

158. *Id*. at 19.

159. *Id*. at 5.

127

Georgia cities provide another example of using mediation in local legislation.[160] Georgia's cities and counties shared revenues and were tasked with devising a distribution plan for those revenues.[161] The parties were given sixty days to negotiate amongst each other and if no agreement was produced, the plan would move to arbitration.[162] With the assistance of mediators, an agreement was finally reached, though not without some strategy on the mediators' part.[163]

Similarly, North Carolina passed a statute allowing the school board to mediate any inadequacies in the school budget with county commissioners.[164] Since the passage of the statute, twenty-nine of the thirty mediations resulted in successful settlements.[165] A particularly interesting byproduct of these mediations was the social bond that was formed during private communications of the mediation.[166] In that situation, lawmakers felt more comfortable to empathize with one another, breaking down adversarial barriers.[167]

This situation references a few of the proposed theories that have led to mediation. For instance, the idea that lawmakers spend less time together because of these technological advancements in traveling, resulting in less of them staying in Washington D.C. and getting to know one another, as referenced in the North Carolina case.[168] In that case, being together in a less formal and more social atmosphere was beneficial to the development of relationships across the aisle and, as a result, the decrease in the adversarial nature of the legislature.[169] Second, the theory that media's overwhelming attention to politics hampers political compromise is also highlighted in the North Carolina case. North Carolina's lawmakers' success in building relationships during these private communications really evidences media's effect in hampering communication.

---

160. *Id.* at 6.

161. *Id.*

162. *Id.*

163. *Id.* "Even more so than private parties, politicians are often suspicious of each other" thus, "mediators used an evaluative mediation style as opposed to a more traditional facilitative style." *Id.* By doing so, "the mediator is more freely able to give advice and suggest equitable resolutions or new avenues of solutions to explore" as opposed to "facilitative mediation in which the mediator does not give advice or make assessments." *Id.*

164. *Id.* at 7.

165. *Id.*

166. *Id.*

167. *Id.* at 8.

168. *Id.* at 7.

169. *Id.* at 8.

128

From looking at both negotiation and mediation strategies in the political legislative realm, several things are evident: (1) they both provide a process to which polarization can be solved, and yet (2) we delay incorporating them into the legislative process because of our democratic norms. Thus, we should reconsider our democratic norms. The founders safeguarded the public from tyranny and over-control through a process of checks and balances, and yet the public is an arm of the government as much as the executive is. Because we have no systematic check, the public must check themselves. A way of doing this would be encouraging the implementation of these tactics in our legislative process.

PART III: CONCLUSION

As America looks forward to the 2016 presidential election, it is apparent that American politics is changing.[170] The rise of candidates such as Bernie Sanders and Donald Trump evidences the growing support for extremists.[171] Both candidates play on the growing anger and fear that the public has with the government.[172] Thus, at the moment, "the angry voices are the loudest in the room."[173]

Clearly, national anger is misplaced. If the American people are angry and fearful of our government's lack of productivity or action in certain policy issues, extremism is not the answer. Divided legislatures are not productive, and chances are, with an extreme president, the American government will continue along an unproductive path. Mediation and negotiation tactics would provide a process for legislatures to overcome this division. For the sake of securing a productive and efficient democracy, we must encourage the implementation of these tactics for "[t]he American spirit wears no political label. In service to others and yes, in sacrifice for our country, there are no Republicans; there are no Democrats; there are only Americans."[174]

---

170. Peter W. Stevenson, *What Unities Bernie Sanders and Donald Trump? Anger.*, THE WASHINGTON POST (Jan. 26, 2016), https://www.washingtonpost.com/news/the-fix/wp/2016/01/26/what-bernie-sanders-and-donald-trump-have-in-common-theyre-angry-and-proud-of-it/.

171. *Id.*

172. *Id.*

173. *Id.*

174. John Kerry, *Commencement Address at Bedford High School in Toledo, Ohio*, The American Presidency Project (June 6, 2004), http://www.presidency.ucsb.edu/ws/?pid=29699.

129

130

Texas A&M University School of Law
Texas A&M Law Scholarship

Faculty Scholarship

2015

Transforming News: How Mediation Principles Can Depolarize Public Talk

Carol Pauli
Texas A&M University School of Law, carol.pauli@law.tamu.edu

Recommended Citation

Transforming News: How Mediation Principles Can Depolarize Public Talk

Carol Pauli*

ABSTRACT

News media interviews bring opposing voices into the public forum where, ideally, audience members can deliberate and reach democratic compromise. But in today's politically polarized atmosphere, partisans increasingly accuse each other of being a threat to the country, and prospects for compromise have suffered. Journalists have been urged to take a more affirmative role, promoting problem solving and opposing conflict. They have stopped short, citing professional norms that demand a stance of neutral detachment.

This article turns to the principles of transformative mediation. Like journalism, it is detached from any goal of settlement. It aims instead at increasing the capacity of participants to clarify their views and respond with generosity to the views of opponents. This is a goal that journalism can embrace and the public forum can use. This article draws on empirical research and offers practical suggestions, using recent news interviews to illustrate both problems and potential directions.

I. INTRODUCTION

Journalists are third-party neutrals with a freewheeling ability to seek out conflicts and bring the voices of opposing parties into the public sphere.

Here, the news media are supposed to create a forum for the kind of public compromise[1] considered essential to democracy.[2] But people on both the right and left rank journalists among the lowest occupations in terms of contribution to society,[3] and the public forum has grown so polarized that politically active Americans increasingly see their opponents' policies as, not just wrong, but "so misguided that they threaten the nation's well-being."[4] Ordinary people appear to be tuning out.[5]

This article proposes that mediation insights can help journalists improve and develop their contribution to a functioning public forum. Already, some news anchors have famously employed the moves of

---

1. BILL KOVACH & TOM ROSENSTIEL, THE ELEMENTS OF JOURNALISM: WHAT NEWSPEOPLE SHOULD KNOW AND THE PUBLIC SHOULD EXPECT 166 (2007) ("This is the sixth principle or duty of the press: Journalism must provide a forum for public criticism and compromise.").

2. Jurgen Habermas, *Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research*, 16 COMM. THEORY 411, 412 (2006). "An essential element of a liberal democracy is that the mass media are diverse and independent and give mass audiences access to the consideration of public opinion." *Id.*

3. Pew Research Center, *Public Esteem for Military Still High*, PEW RES. RELIGION & PUB. LIFE PROJECT (July 11, 2013), http://www.pewforum.org/2013/07/11/public-esteem-for-military-still-high/ (noting that, in a survey ranking ten occupations in public esteem, journalists outranked only business executives and lawyers).

4. PEW RESEARCH CENTER, POLITICAL POLARIZATION IN THE AMERICAN PUBLIC 11 (2014), *available at* http://www.people-press.org/files/2014/06/6-12-2014-Political-Polarization-Release.pdf.

5. Markus Prior, *Media and Political Polarization*, 16 ANN. REV. POLIT. SCI 101, 109 (2013).

mediators in on-air interviews; they have brought disputing parties together, used questions to clarify issues, and invited the parties to generate solutions.[6] These steps are among the hallmarks of the most widely used form of mediation, called the interest-based, facilitative, or problem-solving model.[7] For journalists, however, conducting this kind of interview is problematic because it threatens to draw them into the content of the issues and away from their ideal position of neutral detachment. Journalists are ambivalent, at best, about engaging in problem solving, and they are inclined to view such moments in interviews as professional lapses, perhaps even cause for embarrassment.[8]

Therefore, this article directs its attention to a less widely used approach to conflict resolution: transformative mediation and allied relational models for facilitating public disputes. On the surface, transformative mediation has some intriguing similarities to news interviews; both highlight the areas of disagreement in a conflict, and both unabashedly and precisely repeat even the hurt and hostile words of the disputing parties.[9] More importantly, the transformative model maintains a detachment from the issues that are in dispute, and it pointedly makes no attempt to resolve them.[10] Instead its

---

6. Eytan Gilboa, *Media-Broker Diplomacy: When Journalists Become Mediators,* 22 CRITICAL STUD. IN MEDIA COMM. 99, 102 (2009) (retracing several examples, including perhaps the best known: CBS News anchor Walter Cronkite's broadcast interviews in 1977 with Egyptian President Anwar Sadat and his enemy, Israeli President Menachem Begin, which led to Sadat's trip to Jerusalem the following week and the peace accord between the two countries); *see also* Carol Pauli, *News Media as Mediators,* 8 CARDOZO J. CONFLICT RESOL. 717 (2007).

7. *See, e.g.,* CARRIE J. MENKEL-MEADOW ET AL., DISPUTE RESOLUTION: BEYOND THE ADVERSARIAL MODEL 266-67 (2005).

8. Gilboa, *supra* note 6, at 103 (noting that after prompting the dialogue that led to the Camp David accords, "Cronkite refused to take credit for his mediation role").

9. Robert A. Baruch Bush & Joseph P. Folger, *Transformative Mediation: Core Practices, in* TRANSFORMATIVE MEDIATION: A SOURCEBOOK 31, 39 (Joseph P. Folger et al. eds., 2010).

10. ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: THE TRANSFORMATIVE APPROACH TO CONFLICT 65-66 (2005) ("[D]efinition[s] of mediation itself, and the mediator's role, in the transformative model . . . differ markedly from the normal definitions found in training materials and practice literature—in which mediation is usually defined as a

goal, through incremental moves, is to increase the capacity of the parties to
clarify and articulate their own views and to hear each other's perspectives,
thus improving their ability to work out their problem on their own.[11]

This goal is consistent with the ideal public forum, in which citizens
explain the reasons for their positions and listen to competing ideas.[12]
Because mediators in the transformative model do not aim at settlement,
their role is also consistent with the detachment of reporters. Still, importing
transformative mediation into a news context is difficult; news interviews
have a public audience, which is something mediators generally avoid—
even when addressing public issues.[13]

The presence of the public audience both limits and expands the
potential use of the transformative approach. In front of cameras or
microphones, partisan interviewees are unlikely to suddenly discover a new
appreciation of each other's perspectives; however, audience members,
observing the interview unseen, are not so constrained. They are free to
experience small, spontaneous moments of greater personal clarity and
deeper understanding. So, while the transformative model, if employed in
news interviews, might possibly help partisan interviewees hear each other,

---

process in which a neutral third party helps the parties to reach a mutually acceptable resolution . . . .
By contrast, in the transformative model . . . [t]he mediator's role is to help the parties make positive
interactional shifts . . . . No mediator can "get" parties to shift out of weakness or self-absorption,
nor should he [sic] try."); *Institute for the Study of Conflict Transformation, Inc., Mediation:
Principles & Practice, in* MEDIATION CENTER OF DUTCHESS COUNTY, BASIC TRANSFORMATIVE
MEDIATION MANUAL 23 (2010) ("Settlement is *not* the *mediator's* goal.").

11. Robert A. Baruch Bush & Joseph P. Folger, *Transformative Mediation: Theoretical
Foundations, in* TRANSFORMATIVE MEDIATION: A SOURCEBOOK 15, 25-26 (Joseph P. Folger et al.
eds., 2010).

12. DIANA C. MUTZ, HEARING THE OTHER SIDE: DELIBERATIVE V. PARTICIPATORY
DEMOCRACY 125 (2006) ("We want the democratic citizen to be . . . aware of all of the rationales for
opposing sides of an issue, yet not to be paralyzed by all of this conflicting information and the
cross-pressures it brings to bear."). Mutz questions, however, how well both of these qualities can
be attained by the same person. *Id.*

13. Richard Chasin et al., *From Diatribe to Dialogue on Divisive Public Issues: Approaches
Drawn from Family Therapy,* 13 MEDIATION Q. 323, 334-35 (1996).

it may be more useful if adapted to increase the capacity of audience members to hear alternate perspectives and clarify their own thinking, prerequisites for engaging more fully in the public forum.

Part II of this article provides an overview of the news media's public forum function. It reviews journalism's longtime underlying hope that accurate reporting of facts can prevent and resolve conflicts. Part III illustrates journalism's ideal of neutral detachment by reviewing its resistance to two recent movements that tried to enlist reporters in problem solving. Part IV outlines the hallmarks of the transformative model of conflict resolution and several variations that it has inspired. All of these models maintain a detachment from the issues and focus, instead, on the relationship of the parties. Part V uses the transformative model to look at the problems and possibilities for journalists interviewing opposing parties in front of an audience. Part VI offers ways that journalists can make use of the insights of transformative mediation to develop and improve their role in fostering a functioning public forum. Part VII suggests areas for further exploration.

II. FAITH IN FACTS

News interviews carry strident voices into the public sphere, where conflicts take shape[14] and people have the opportunity to consider and respond to them. Many Americans learn about the various perspectives that compete for their attention—and their votes—through the voices they encounter in the media. "[M]edia are far more important than interpersonal networks in exposing people to views unlike their own. As a result, the

---

14. MANUEL CASTELLS, COMMUNICATION POWER 301 (2009) ("Social movements are formed by communicating messages of rage or hope. The specific structure of communication of a given society largely shapes social movements. In other words, social movements, and politics, insurgent or not, spring up and live in the public space. Public space is *the space of societal, meaningful interaction where ideas and values are formed, conveyed, supported, and resisted; space that ultimately becomes a training ground for action and reaction.*").

media have the potential to make an extremely important contribution to awareness of diverse political perspectives and thus to national political integration."[15]

To say "the media" is to evoke a wide assortment of models for gathering and disseminating information and opinions. The term may mean anything from propaganda to entertainment.[16] This article focuses on the news media and uses the terms *news* and *journalism* to refer to the fact-based reporting and writing that is described as "the journalism of verification"[17] by Bill Kovach, founder of the Committee of Concerned Journalists,[18] and Tom Rosenstiel, executive director of the American Press Institute. The journalism of verification first aims to produce accurate information and is, in this way, distinguishable from entertainment, "infotainment," and propaganda.[19] Within journalism, this article focuses on

---

15. Diana C. Mutz & Paul S. Martin, *Facilitating Communication Across Lines of Political Difference: The Role of Mass Media*, 95 AM. POL. SCI. REV. 97, 109-10 (2001). "[F]or many Americans [mass media is] the main source of exposure to crosscutting political views. . . . The media send multiple conflicting messages, and in so doing they advance an important aspect of the democratic process." *Id.* at 110.

16. KOVACH & ROSENSTIEL, *supra* note 1, at 173 ("[T]he new Mixed Media Culture of radio and TV talk shows, websites, chat rooms, blogs, bulletin boards and more that now dominate the communications systems has seen the urge to comment replace the need to verify, sometimes even the need to report.").

17. KOVACH & ROSENSTIEL, *supra* note 1, at 78; BILL KOVACH & TOM ROSENSTIEL, BLUR: HOW TO KNOW WHAT'S TRUE IN THE AGE OF INFORMATION OVERLOAD 34 (2010).

18. *Id.* at 79. The Committee of Concerned Journalists began with a 1997 gathering in Cambridge, Mass., of twenty-five editors, broadcasters, and other journalists, worried by falling levels of public trust and fearful that the public was right. The Committee ceased operation at the close of 2011. Its work continues at the Donald W. Reynolds Institute at the Missouri School of Journalism, at the University of Missouri, http://www.rjionline.org/ccj. Kovach and Rosenstiel continue to write on applying the core values of journalism to the changing media landscape. *See, e.g.,* BILL KOVACH & TOM ROSENSTIEL, WARP SPEED: AMERICA IN THE AGE OF MIXED MEDIA (1999).

19. KOVACH & ROSENSTIEL, *supra* note 1, at 79 ("Entertainment—and its cousin 'infotainment'—focuses on what is most diverting. Propaganda selects facts or invents them to serve the real purpose: persuasion and manipulation.").

television and radio news interviews. Although Web-based news delivery is quickly growing,[20] these older, mainstream news platforms still command the largest audiences[21] and supply much of the original news content that is repeated and aggregated by online news sites.

Interviewing has been an essential part of newsgathering since the early 1900s, when journalism began using factual reporting to distinguish itself from an earlier era of opinionated writing.[22] In the 1920s, as journalism began to view itself as a profession, the newly formed American Society of Newspaper Editors adopted "Canons of Journalism," which included the ideals of truthfulness and unbiased accuracy.[23] By the end of the decade, "every major professional society and most major publications had adopted [a code of ethics]."[24] Under their provisions, journalists are to be neutrals[25] who are alert to conflict[26] but do not take sides.[27]

---

20. Mark Jurkowitz, *The Growth in Digital Reporting*, PEW RES. JOURNALISM PROJECT, (Mar. 26, 2014), http://www.journalism.org/2014/03/26/the-growth-in-digital-reporting/.

21. Pew Research Center, *Key Indicators in Media & News*, PEW RES. JOURNALISM PROJECT, (Mar. 26, 2014), http://www.journalism.org/2014/03/26/state-of-the-news-media-2014-key-indicators-in-media-and-news/ (finding that on an average evening in 2013, news programs on the three commercial television networks (ABC, CBS, and NBC) had 22.6 million viewers, and the leading cable networks (Fox, CNN, and MSNBC) combined to draw just under 3 million).

22. MICHAEL SCHUDSON, THE SOCIOLOGY OF NEWS 81-82 (2003).

23. *Id.* at 82.

24. Joseph A. Mirando, *Lessons on Ethics in News Reporting Textbooks, 1867-1997*, 13 J. MASS MEDIA ETHICS 26, 30 (1998) ("The development of tabloids and prohibition-era reporting brought forth a renewed outcry against sensationalism, a revulsion of partisanship and propaganda among newspapers, and an allegiance to objective reporting.").

25. *SPJ Code of Ethics*, SOC'Y OF PROF'L JOURNALISTS, http://www.spj.org/ethicscode.asp (last updated Sept. 6, 2014) ("Journalists should . . . [t]ake responsibility for the accuracy of their work[;] . . . [s]upport the open and civil exchange of views, even views they find to be repugnant[; and] . . . [a]void conflicts of interest, real or perceived.").

26. The "news value" of conflict is pointed out in the opening chapters of many journalism textbooks. *See, e.g.*, MELVIN MENCHER, MELVIN MENCHER'S NEWS REPORTING AND WRITING 60 (10th ed. 2006) ("People and their tribes and their countries have been at war with each other, and with themselves, since history has been kept, and the tales that resulted have been the basis of saga, drama, story and news."); *see also* ANDREW ARNO, ALARMING REPORTS: COMMUNICATING

Kovach and Rosenstiel urge that journalism has a duty to provide a public forum grounded in factual information.[28] In so doing, they reflect a longstanding journalistic hope that shared, verified facts will enable the public to make wise choices among the conflicting views that are inherent in a democracy.[29] Some past leaders in the field have even imagined that factual reporting would lead a worldwide public to global harmony. In the wake of World War I, the general manager of the Associated Press, Kent Cooper, proposed that journalism could rise above politics and nationalism to reveal to the audience a common humanity transcending borders.[30] Cooper argued that "all-inclusive journalism," reporting the daily simple events that filled people's lives, would give readers on both sides of the ocean a sympathetic look at each other,[31] and over his next forty years at the *AP*, he continued to develop what became known as "human interest" news.[32]

CONFLICT IN THE DAILY NEWS 39 (2009) ("[T]he most obvious fact about news is that news is always and only about conflict.").

27.   LARS WILLNAT & DAVID H. WEAVER, THE AMERICAN JOURNALIST IN THE DIGITAL AGE: KEY FINDINGS 11 (2014), *available at* http://news.indiana.edu/releases/iu/2014/05/2013-american-journalist-key-findings.pdf (finding, in a survey of more than 1,000 American journalists, that just over 50% identified themselves as political independents, about 38% as Democrats, and about 7% as Republicans). *But see* Matthew A. Baum & Tim Groeling, *New Media and the Polarization of American Discourse*, 25 POL. COMM. 345, 347-49, 357 (2008) (finding an anti-Republican bias in stories of the Associated Press, in contrast to those of the British wire service, Reuters).

28.   KOVACH & ROSENSTIEL, *supra* note 1, at 167.

29.   For a discussion of conflict in the context of the news media's role in the public forum, see Richard Reuben, *The Impact of News Coverage on Conflict: Toward Greater Understanding*, 93 MARQ. L. REV. 45, 49-51 (2009).

30.   *Journalism in Role for Peace: Institute of Politics Hears of International Project*, L.A. TIMES, Aug. 8, 1926, at 7, *available at* ProQuest Historical Newspapers: Los Angeles Times (1881-1989).

31.   *Id.*

32.   Malcolm Stephenson, *Kent Cooper, Former Chief of Associated Press, Dies*, EVENING NEWS   (Newburgh,   N.Y.),   Feb.   1,   1965,   at   7A,   *available   at* http://news.google.com/newspapers?nid=1982&dat=19650201&id=hvVfAAAAIBAJ&sjid=1m0NA AAAIBAJ&pg=2661,50397.

After World War II, the Commission on Freedom of the Press—which
came to be known as the Hutchins Commission[33]—also expressed faith that
factual journalism could prevent violent conflict.   The Commission
presented the task with urgency, summoning up the fresh memory of Nazi
propaganda and the U.S. atomic bombing of Japan

> With the means of self-destruction that are now at their disposal, men must live, if they
> are to live at all, by self-restraint, moderation, and mutual understanding. They get their
> picture of one another through the press. The press can be inflammatory, sensational, and
> irresponsible .... On the other hand, the press ... can help create a world community by
> giving men everywhere knowledge of the world and of one another.[34]

Even now, in the face of continuing conflict, war reporters still express
some fragile faith that accurate reporting can help to end violence.   Two
years before she was killed in a shelling attack in Homs, Syria,[35] war
correspondent Marie Colvin referred to this hope to explain why she risked
her life:[36]

> [S]omeone has to go there and see what is happening. You can't get that information
> without going to places where people are being shot at, and others are shooting at you.
> The real difficulty is having enough faith in humanity to believe that enough people—be

---

    33.   The Commission was initiated by Henry R. Luce, a founder of Time, Inc., and funded by
Time, Inc., and Encyclopedia Britannica, Inc. University of Chicago Chancellor Robert M. Hutchins
and Harvard Law Professor Zechariah Chaffee, Jr. served as chairman and vice chairman,
respectively. COMMISSION ON FREEDOM OF THE PRESS, A FREE AND RESPONSIBLE PRESS, ii, v
(1947), *available at* http://archive.org/stream/freeandresponsib029216mbp#page/n21/mode/2up
10/11/13. The Commission's emphasis on the responsible use of communication helped to shape the
"social responsibility" theory of the press. FRED S. SIEBERT ET AL., FOUR THEORIES OF THE PRESS
75 (1974).
    34.   COMMISSION ON FREEDOM OF THE PRESS, *supra* note 33, at 4.
    35.   John Swain, *Marie Colvin: The Last Assignment,* SUNDAY TIMES (London), Feb. 26, 2012,
at 2.
    36.   M. David Arant & Philip Meyer, *Public and Traditional Journalism: A Shift in Values?*
13 J. MASS MEDIA ETHICS 205, 209-10 (1998).

they government, military or the man on the street—will care when your file reaches the printed page, the website, or the TV screen.[37]

Factual reporting ideally supplies Americans with an information pool that opposing partisans can share as they deliberate.[38] Facts can interrupt the information and reputational cascades that encourage people to shift toward extreme points of view.[39] Yet research in political science, psychology, and media studies finds that, even when verified facts are available, emotions play a large—perhaps controlling—role in civic decision making.[40] Studies have found that angry people are drawn to facts that maintain or exacerbate their anger, and people who feel self-oriented anxiety are drawn to facts related to arousal and self-protection.[41] People choose which news outlets to expose themselves to,[42] and this choice enables them to limit the range of

---

37. Marie Colvin, Speech at St. Bride's Church Service in Memory of Journalist Casualties (Nov. 10, 2010), *available at* http://www.theguardian.com/commentisfree/2012/feb/22/marie-colvin-our-mission-is-to-speak-truth.

38. BILL KOVACH & TOM ROSENSTIEL, BLUR: HOW TO KNOW WHAT'S TRUE IN THE AGE OF INFORMATION OVERLOAD 180 (2010) ("A community's news institutions, new or old, can serve as public squares where we citizens can monitor voices from all sides, not just those in our own ideological affinity group.").

39. *See* Cass Sunstein, *Essay: Deliberative Trouble? Why Groups Polarize*, 110 YALE L.J. 71, 83 (2000).

40. George E. Marcus, *Emotions in Politics*, 3 ANN. REV. POL. SCI. 221, 231 (2000) (summarizing research in political science and psychology); Kang Namkoong et al., *The Politics of Emotion: News Media, Attention, Emotional Responses, and Participation During the 2004 U.S. Presidential Election*, 15 MASS COMM. & SOC'Y 25, 27 (2012) (referring to research in political science and media studies).

41. Cynthia A. Hoffner et al., *Why We Watch: Factors Affecting Exposure to Tragic Television News*, 12 MASS COMM. & SOC'Y 193, 209 (2009) (surveying more than 300 undergraduate students in the wake of the terrorist attack of September 11, 2001, and finding that those who scored higher in personal distress said that they watched news stories less out of empathy than out of curiosity about morbid events—a motive associated with a desire for arousal and maybe self-protection).

42. *See, e.g.*, Shanto Iyengar & Kyu S. Hahn, *Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use*, 59 J. COMM. 19, 33-35 (2009).

facts that enter their information worlds.[43] When they encounter identical facts, people with different political leanings tend to make different interpretations.[44] The mere exploration of a public issue can polarize audience members.[45] What is even worse for factual reporters and democratic compromise is that increased news coverage of an issue appears to widen the gap between partisans.[46]

III. CONFLICT OVER RESOLUTION

Two recent movements have urged journalism to move beyond factual reporting and invest in problem solving. The civic journalism movement has called on journalists to actively help audience members address civic problems together in their communities.[47] The peace journalism movement has urged them to promote peace efforts in violent conflict. In both cases, traditional journalism has stopped short, unable to find a way forward that would not sacrifice dispassionate reporting. A brief recounting of the resistance to these movements illustrates journalism's discomfort with playing a problem-solving role.

---

43. Amy Mitchell et al., *Political Polarization & Media Habits*, PEW RES. JOURNALISM PROJECT (Oct. 21, 2014), http://www.journalism.org/2014/10/21/political-polarization-media-habits/ (finding that people with consistently conservative views overwhelmingly use and trust news outlets different from the ones used and trusted by people with consistently liberal views).

44. Matthew Gentzkow & Jesse M. Shapiro, *Ideological Segregation Online and Offline*, 126 Q. J. ECON. 1799, 1832 (2011); Prior, *supra* note 5, at 109.

45. Sunstein, *supra* note 39, at 80.

46. Jennifer Jerit & Jason Barabas, *Partisan Perceptual Bias and the Information Environment*, 74 J. POL. 672, 682 (2011) ("People are motivated to see the world in a manner that is consistent with their political views. This results in a selective pattern of learning in which partisans have higher levels of knowledge for facts that confirm their world view and lower levels of knowledge for facts that challenge them. This basic pattern is exaggerated on topics receiving extensive news coverage.").

47. Jay Rosen, *Questions and Answers about Public Journalism*, 1 JOURNALISM STUD. 679, 680 (2000).

A. *Civic Journalism*

In the late 1980s, New York University Professor Jay Rosen observed that, unlike the public of democratic theory, the real public was inattentive, splintered on most issues, and unable to function.[48]   He proposed that journalism should promote civic life by thinking of audience members as citizens and re-engaging them in public discussions.[49] His approach became known as "civic" or "public" journalism.[50]

Rosen first brought his idea to the Associated Press Managing Editors convention.[51]   At that time, even before competition from the Internet, newspapers were facing declines in readership that happened to parallel declines in membership in many civic organizations.[52]   Rosen's approach encouraged newspapers to incorporate the views of citizen-readers in their reporting,[53] to link interested citizens to each other,[54] and even to gather citizens for meetings and dialogue.[55]

In 1996, James Fallows, then Washington Editor of the *Atlantic Monthly*, recommended civic journalism practices.[56]   He wrote that journalism was failing democracy because it had grown "arrogant, cynical, scandal-minded, and destructive."[57]   The watchdog role of the press, continually criticizing government, he wrote, had led to a public sense of

---

48. JAY ROSEN, WHAT ARE JOURNALISTS FOR? 19 (1999).

49. Rosen, *supra* note 47, at 679-80.

50. ROSEN, *supra* note 48, at 21.

51. *Id.* at 19.

52. *Id.*

53. Burton St. John III, *Newspapers' Struggles with Civic Engagement: The U.S. Press and the Rejection of Public Journalism as Propagandistic*, 10 COMM. REV. 249, 256 (2007).

54. Rosen, *supra* note 47, at 680.

55. Albert W. Dzur, *Public Journalism and Deliberative Democracy*, 34 POLITY 313, 316 (citing ARTHUR CHARITY, DOING PUBLIC JOURNALISM 19 (1995)).

56. JAMES FALLOWS, BREAKING THE NEWS: HOW THE MEDIA UNDERMINE AMERICAN DEMOCRACY 247 (1996).

57. *Id.* at 3.

mistrust, hopelessness, and disengagement,[58] not to mention a declining image of the news media.[59]   Fallows quoted a study that found that communities with healthy civic engagement understood that issues were entangled with each other.[60] They valued ambivalence on thorny questions, and they had local institutions that helped them connect to each other.[61]

Some newspapers, especially small- to medium-circulation papers, responded to the call of civic journalism.   The *Charlotte Observer* interviewed residents of high-crime areas and asked them to suggest solutions.[62] The *Columbus Ledger-Enquirer* surveyed readers and wrote an eight-part series about the future of the Georgia city, including its economy, infrastructure, and schools—and when no public action resulted, the newspaper took the next step.[63] It offered a venue and helped run a meeting, which led to more meetings, discussions, and task forces.[64]

Many other newspapers were critical. The *New York Times*, responding to Fallows's book in an editorial, said it was dangerous to have reporters and editors become "public policy missionaries."[65]   Others worried about newspaper involvement in civic projects. "I think what Columbus did was bad," said Howard Schneider, managing editor of *Newsday.*[66] He worried that, by getting involved and making the community agenda its own, the newspaper had weakened its future ability to report on harsh realities.[67] Some critics argued that engagement in the community blurred the line

---

58.   *Id.* at 240-43.

59.   Dzur, *supra* note 55, at 315.

60.   FALLOWS, *supra* note 56, at 244-45.

61.   *Id.* at 244-46.

62.   Dzur, *supra* note 55, at 315.

63.   ROSEN, *supra* note 48, at 13.

64.   *Id.* at 29-30.

65.   Howell Raines, *Editorial Notebook: The Fallows Fallacy; A Plan for Turning Reporters Into Lackeys*, N.Y. TIMES, Feb. 25, 1996, http://www.nytimes.com/1996/02/25/opinion/editorial-notebook-the-fallows-fallacy-a-plan-for-turning-reporters-into-lackeys.html.

66.   ROSEN, *supra* note 48, at 31.

67.   *Id.*

between reporting and advocacy.[68] For journalists, engaging in advocacy
was taboo.[69] Others cautioned that public journalism went too far, assuming
jobs that belonged to government[70] or force-feeding the public its own
concept of citizenship.[71] "Many journalists suspected that the proposed
abandonment of objectivity and remoteness for a facilitative role was an
attempt to maneuver the public's attention and action in a way that could
benefit privileged interests."[72]

In 1998, a survey of 375 newspapers found that, while most journalists
wanted to help their communities, they did not favor doing so by actively
engaging with civic groups or by organizing programs.[73] Instead, and
overwhelmingly, they gave the strongest support to their traditional roles of
exposing government wrongdoing and providing practical information to the
public.[74] By the year 2000, more than 300 newspapers, one-fifth of

---

68. Tanni Haas & Linda Steiner, *Public Journalism: A Reply to Critics*, 7 JOURNALISM 238,
248 (2006).

69. *See, e.g.*, N.Y. TIMES, ETHICAL JOURNALISM: A HANDBOOK OF VALUES AND PRACTICES
FOR   THE   NEWS   AND   EDITORIAL   DEPARTMENTS   (2004)   *available   at*
http://nytco.com/pdf/NYT_Ethical_Journalism_0904.pdf ("Staff members may not serve on
government boards or commissions, paid or unpaid. They may not join boards of trustees, advisory
committees or similar groups . . . . Normally the restriction will not apply to organizations that are
highly unlikely to generate news of interest to the Times and that do not generally seek to shape
public   policy.");   AP   News   Values   &   Principles,   ASSOCIATED   PRESS,
http://www.ap.org/company/news-values (last visited Jan. 11, 2015) ("Editorial employees are
expected to be scrupulous in avoiding any political activity, whether they cover politics regularly or
not.").

70. Dzur, *supra* note 55, at 315.

71. Will Friedman, *Coming to Public Judgment: Strengthening Impacts, Exploring National
Possibilities, in* TOWARD WISER PUBLIC JUDGMENT 141 (Daniel Yankelovich & Will Friedman eds.,
2011) (quoting Tom Goldstein, *Good Question* (review of JAY ROSEN, WHAT ARE JOURNALISTS
FOR?), N.Y. TIMES, Nov. 14, 1999).

72. St. John, *supra* note 53, at 257.

73. Arant & Meyer, *supra* note 36, at 216.

74. *Id.* at 216-17.

newspapers in the United States, practiced civic journalism in some form,[75] but even among these, more than half specified that their goal was in line with traditional journalism values: to inform and raise awareness.[76] Today, journalists rank the traditional watchdog role as their most important role, with a higher percentage seeing it as "extremely important" than even during the Vietnam era.[77]

B. Peace Journalism

As some journalists were experimenting with civic journalism in their communities, others were recommending a similar affirmative and engaged role on battlefields. After the Persian Gulf War in 1991, correspondents Annabel McGoldrick and Jake Lynch began urging their colleagues to go beyond fact-finding to help society consider non-violent approaches to conflict.[78] Their ideas were based on the work of sociologist Johan Galtung, who had named this approach "peace journalism."[79] He accused "war and violence journalism"[80] of highlighting the immediacy and drama of battle while ignoring the slower process of negotiation.[81] Rather than focusing on a conflict as a zero-sum game, he advocated covering peace talks, exploring possibilities for win-win solutions, and aiming at preventing violence.[82]

---

75.   LEWIS A. FRIEDLAND & SANDY NICHOLS, MEASURING CIVIC JOURNALISM'S PROGRESS: A REPORT   ACROSS   A   DECADE   OF   ACTIVITY   18   (2002),   *available   at* https://web.archive.org/web/20130724032934/http://www.pewcenter.org/doingcj/research/measurin gcj.pdf

76.   *Id.*

77.   WILLNAT & WEAVER, *supra* note 27, at 14 (comparing 78.2% in 2013 to 76% in 1971).

78.   Thomas Hanitzsh, *Journalists as Peacekeeping Force? Peace Journalism and Mass Communication Theory*, 5 JOURNALISM STUD. 483, 484 (2004).

79.   Johan Galtung, *Peace Journalism*, 30 MEDIA ASIA 177, 177 (2003).

80.   Johan Galtung, *Peace Journalism: What, Why, Who, How, When, Where?*, *in* THE PEACE JOURNALISM CONTROVERSY 20 (Wilhelm Kempf ed., 2008).

81.   *See id.* at 22.

82.   Galtung, *supra* note 79, at 178.

Galtung advised this approach, not only for covering battlefields, but also for reporting on conflicts at all levels, from national strife to domestic violence and child abuse.[83]

Along similar lines, former British Broadcasting Corporation (BBC) foreign affairs correspondent Martin Bell has advocated an ethical duty for journalists to side with the victims of war.[84] He recommends that reporters practice a "journalism of attachment" to the conflicts they cover by engaging in a moral enterprise that "will not stand neutrally between good and evil, right and wrong, the victim and the oppressor."[85] However, peace journalism and the journalism of attachment, like civic journalism, have been met by skepticism from traditionalists.[86] Among the most vocal opponents of peace journalism, BBC reporter David Loyn has called its prescriptions "dangerous."[87] Loyn insists that to encourage peacemaking is not the reporter's business.[88] Peace journalism has been more easily adopted in developing and post-conflict nations than amid the established—and professionally detached—journalism traditions of Washington and London.[89]

Although these two movements did not revolutionize journalism, their emergence reveals some concern and frustration within the field at the

---

83. *Id.* at 179.

84. Wilhelm Kempf, *Constructive Conflict Coverage: A Social-Psychological Model, in* THE PEACE JOURNALISM CONTROVERSY 34 (Wilhelm Kempf ed., 2008).

85. Karoline Von Oppen, *Reporting from Bosnia: Reconceptualising the Notion of a 'Journalism of Attachment,'* 17 J. CONTEMP. EUR. STUD. 21, 24 (quoting Martin Bell, *TV News: How Far Should We Go?*, 8 BRIT. JOURNALISM REV. 7 (1997)).

86. Reuben, *supra* note 29, at 83-84.

87. David Loyn, *Good Journalism or Peace Journalism?, in* THE PEACE JOURNALISM CONTROVERSY 53 (Wilhelm Kempf ed., 2008).

88. *Id.* at 54. For a response to such criticism, see Jake Lynch, *Peace Journalism and its Discontents*, 6 CONFLICT & COMM. ONLINE, no. 2, 2007, *available at* http://www.cco.regener-online.de/2007_2/pdf/lynch.pdf.

89. *See* Robert A. Hackett, *Journalism for Peace and Justice: Towards a Comparative Analysis of Media Paradigms*, 4 STUD. SOC. JUST. 179, 186 (2010).

results of traditional factual reporting. The movements' failure to win widespread acceptance leaves open the question of how journalists might still further the functioning of the public forum without violating their own professional norms. Transformative mediation suggests a possible approach by offering two propositions: that what participants in a productive forum need, in addition to facts, is the capacity to hear each other and clarify their own ideas[90]—and that third-party neutrals can support growth in that capacity.

IV. TRANSFORMATIVE MEDIATION

The transformative model of mediation shares a number of characteristics with other forms of mediation. Similar to problem-solving mediation, the transformative model allows opposing parties to discover and consider issues that might hold the key to resolving their conflict, even when those issues have no legal remedy; it also offers parties an informal, consensual process, unlike the win-lose arena of the courtroom.[91] The transformative model also gives control of the outcome to the parties; they shape the settlement agreement, if there is one, and no authority imposes it.[92]

The transformative model, however, seeks a goal that is different from that of the problem-solving model. Problem-solving mediation aims to help parties move away from their fixed positions in order to discover their

---

90.   *See* Hiro Aragaki, *Deliberative Democracy as Dispute Resolution? Conflict, Interests, and Reasons*, 24 OHIO STATE J. ON DISP. RESOL. 406, 431-35 (2009) (detailing how relational conflict undermines the ability of participants in a deliberative democracy to reason with each other over substantive issues); *see also* Steven A. Rosell & Heidi Gantwerk, *Moving Beyond Polls and Focus Groups*, *in* TOWARD WISER PUBLIC JUDGMENT 110-15 (Daniel Yankelovich & Will Friedman eds., 2011) (proposing that empathetic listening is essential to working through democratic problem solving), and Carrie Menkel-Meadow, *Deliberative Democracy and Conflict Resolution*, DISP. RESOL. MAG., Winter 2006, at 18, 20 (discussing the role of empathy in both deliberative democracy and conflict resolution).
91.   BUSH & FOLGER, *supra* note 10, at 8.
92.   *Id.*

underlying interests, and it helps generate creative ways in which both might
satisfy those interests.[93] Transformative mediation does not focus on sorting
out the parties' problems, discovering their interests, or generating
solutions.[94] Instead, it aims at improving the quality of their interaction by
changing the way the parties understand themselves and each other.[95]

As formulated by Robert A. Baruch Bush and Joseph P. Folger, scholars
in law and communication respectively,[96] transformative mediation holds
that people in conflict experience "a loss of personal strength and clarity."[97]
In response, they become self-absorbed and accusatory.[98] Behaving in ways
that are increasingly hostile, they prompt reciprocal responses from each
other, in a continuing downward spiral.[99] Transformative mediation aims to
transform individuals from fearful, defensive, or self-centered beings into
confident, responsive, and caring ones—who are then quite capable of
grappling with the issues by themselves.[100]

---

93. *See e.g.*, Lela P. Love & Kimberlee K. Kovach, *ADR: An Eclectic Array of Processes,
Rather Than One Eclectic Process*, 2000 J. DISPUTE RESOL. 295, 317 (2000).

94. CARRIE J. MENKEL-MEADOW ET AL., DISPUTE RESOLUTION: BEYOND THE ADVERSARIAL
MODEL 309 (2005).

95. BUSH & FOLGER, *supra* note 10, at 18.

96. Bush is the Harry H. Rains Distinguished Professor of Arbitration and Alternative Dispute
Settlement Law at Hofstra University's Maurice A. Dean School of Law. HOFSTRA LAW,
http://law.hofstra.edu/directory/faculty/fulltime/baruch/ (last visited Jan. 25, 2015). Folger holds a
Ph.D. in Communication Sciences and teaches conflict processes and mediation as a professor in the
College of Education at Temple University. TEMPLE UNIVERSITY COLLEGE OF EDUCATION,
http://education.temple.edu/faculty/joseph-p-folger-phd (last visited Jan. 25, 2015).

97. Joseph P. Folger, *Harmony and Transformative Mediation Practice: Sustaining
ideological Differences in Purpose and Practice*, 84 N.D. L. Rev. 823, 840 (2008).

98. *Id.* at 841.

99. Dorothy Della Noce & Hugo C. M. Prein, *The Case for Transformation: A Review of
Theoretical and Empirical Support, in* TRANSFORMATIVE MEDIATION: A SOURCEBOOK 17-18
(Joseph P. Folger et al. eds., 2010).

100. Geoffrey A. Drucker, *The Postal Service's Decision to Use Transformative Mediation, in*
RESOURCE BOOK FOR MANAGING EMPLOYMENT DISPUTES 5 (2004) (recounting the adoption of
transformative mediation by the U.S. Postal Service "REDRESS" program and observing, "While

To achieve this, the transformative mediator tries to support each party in making two changes: first, the mediator supports the parties in clarifying their own needs and strengths, and second, the mediator supports them in empathizing with each other. Transformative mediation calls these twin goals *empowerment* and *recognition*.[101] The meanings of these two key words should be distinguished from their meanings in other contexts. *Empowerment* does not refer to gaining authority, but to gaining personal clarity and strength.[102] *Recognition* does not refer to identifying something that has been seen before, but to "letting go—however briefly or partially— of one's focus on self and becoming interested in the perspective of the other party as such, concerned about the situation of the other as a fellow human being, not as an instrument for fulfilling one's own needs."[103] Recognition is also characterized as "interpersonal understanding and compassion."[104]

Transformative mediation attempts to support any "shift" in the internal responses of the parties toward empowerment and recognition, and it uses some techniques that are starkly different from those of problem-solving mediation. Where the problem-solving mediator will often highlight areas of agreement between the parties in order to help build momentum toward an overall settlement,[105] the transformative mediator highlights areas of

---

the focus of transformative mediation is on the parties' relationship, settlement of the dispute that brought them to the table is a natural and common outgrowth of the process").

101. BUSH & FOLGER, *supra* note 10, at 18.

102. *Id.* at 76-77 ("[S]upporting [party] empowerment does not mean adding to the strength of either party by becoming an advocate, adviser or counselor.... [It] does not require—or involve at all—the mediator's taking sides, expressing judgments, or being directive, all of which are central aspects of advice giving and advocacy.").

103. *Id.* at 77.

104. Joseph P. Folger & Robert A. Baruch Bush, *Transformative Mediation and Third-Party Intervention: Ten Hallmarks of a Transformative Approach to Practice*, 13 MEDIATION Q. 263, 275 (1996).

105. Lela P. Love & Joseph B. Stuhlberg, *Targets and Techniques to Generate Movement, in* CARRIE J. MENKEL-MEADOW ET AL., DISPUTE RESOLUTION: BEYOND THE ADVERSARIAL MODEL 347, 348 (2005).

disagreement.[106]   Where the problem-solving mediator might rephrase inflammatory words from a disputing party in order to turn them into reasonable issues for discussion,[107] the transformative mediator does not.[108] The transformative mediator sees hostile words, sarcasm, and haranguing as signs that a party feels weak.[109]   In response, the transformative mediator will echo the troublesome statements immediately and precisely, purposely matching even the excited energy of the speaker.[110]   The technique is called "reflection."[111]

Reflection is thought to work in at least three ways. First, it "allows parties to more deeply consider the implications of their own remarks . . . and to restate and refine comments that they may have previously offered."[112]   Second, an accurate reflection reassures a party that he or she has been heard.[113]   This reassurance offers an opportunity for the party to feel an increased sense of empowerment, even if only incrementally.

---

106. *Institute for the Study of Conflict Transformation, supra* note 10, at 70.

107. *See, e.g.,* Lela Love, *Deconstructing Dialogue and Constructing Understanding, Agendas, and Agreements,* 38 FAM. & CONCILIATION COURTS REV. 27, 33 (2000) ("Both recipients of and witnesses to put-downs and insults . . . tend naturally to react with alarm, heightened adrenaline and (especially for the recipient) an attack response. Mediators must hear and, at appropriate points, reframe the feelings that generate such statements.").

108. Folger & Bush, *supra* note 104, at 272 ("[I]nstead of treating emotion as static to be vented and removed, the transformative mediator considers emotion as a rich form of expression that, when unpacked and understood, can reveal plentiful information about the parties' views of their situation and each other—information that can then be used to foster both empowerment and recognition.").

109. Problem-solving mediators also recognize that venting emotions can further a dialogue. *See, e.g.,* Love, *supra* note 107, at 33 ("It is often the case that both parties to a conflict have similar feelings of anger and frustration . . . . [S]ometimes the venting and acknowledgement alone can shift the feelings themselves. Mediators should be trained to hear insults and think 'this person is/may be upset.'").

110. Folger, *supra* note 97, at 823; *see also Institute for the Study of Conflict Transformation, supra* note 10, at 45.

111. *Institute for the Study of Conflict Resolution, supra* note 10, at 42.

112. Folger, *supra* note 97, at 823.

113. *Institute for the Study of Conflict Resolution, supra* note 10, at 47.

Finally, reflection may also allow the other party to hear the hostile statement in a new way, now that it has been pronounced in the different, and presumably less threatening, voice of the mediator.[114]

Transformative mediators also offer occasional "summaries," in which they recount the subjects brought up by the two parties and highlight the topics of disagreement.[115]   Transformative mediators do not rephrase, reframe, or prioritize these issues.   Instead, they offer the summary as an opportunity for parties to take stock before the parties take their conversation in its next direction.[116] The summary is thought to further empowerment by letting parties see clearly an entire segment of their conversation.[117]   It is thought to further recognition by reminding parties of each other's different views."[118]

Transformative   mediation   cites   theoretical   underpinnings   in psychology.[119] The clinical psychologist Carl Rogers used "reflecting" as a tool in nondirective counseling.[120] He described reflection as, ideally, "a clear mirror image of the meanings and perceptions that make up [a client's] world of the moment—an image that is clarifying and insight producing."[121] Rogers wrote that even a client's negative expressions—if accepted and recognized by the therapist—were predictably followed by expressions of love and social connection.[122] Rogers also offered summaries to sharpen and clarify differences.[123]   The downward spiral, in which hostility undermines confidence and trust, is also described by social psychologist Morton

---

114. *Id.*
115. *Id.* at 48.
116. *Id.* at 48-49.
117. *Id.* at 50.
118. *Id.*
119. Della Noce & Prein, *supra* note 99, at 105-17.
120. *Id.* at 109.
121. *Id.*
122. *Id.* at 107-08.
123. *Id.* at 109-10.

Deutsch[124] who has found that parties in conflict respond to each other reciprocally—or, as very briefly summarized, "cooperation breeds cooperation, while competition breeds competition."[125]

The transformative model has been employed beyond intimate mediation sessions, and its proponents claim the potential to address ethnic, political,[126] and other large-scale societal problems.[127] Some, including Judith Saul and Scott Sears, have brought transformative principles to broad community discussions of public policy[128] and to multi-party mediations.[129] To maintain their transformative orientation in these larger-scale disputes,

---

124. *Id.* at 114.

125. MORTON DEUTSCH, THE RESOLUTION OF CONFLICT 367 (1973). Deutsch cautions that this phrase is too condensed, after explaining that "the strategy of power and the tactics of coercion, threat, and deception result from, and also result in, a competitive relationship. Similarly, the strategy of mutual problem solving and the tactics of persuasion, openness, and mutual enhancement elicit, and also are elicited by, a cooperative orientation." *Id.* at 365.

126. Joseph P. Folger & Robert A. Baruch Bush, *Transformative Practice in Ethno-Political Conflict: An Emerging Initiative, in* TRANSFORMATIVE MEDIATION: A SOURCEBOOK 417-21 (Joseph P. Folger et al. eds., 2010).

127. ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: RESPONDING TO CONFLICT THROUGH EMPOWERMENT AND RECOGNITION 46-53 (1994). In this first edition of their book, the authors, recalled news coverage of the Rodney King case, and expressed concern that the dispute resolution field had not responded. They asked what it might have done— or might do in the future—to "reduce the sense of alienation and division that . . . probably fueled the riots." *Id.* at 48. They contrasted this with the optimism of the early mediation movement of the late 1960s, which had expected "to change the conditions that fueled the disorders of that decade, even if only gradually." *Id.* While, in their view, the problem-solving model had retreated from grappling with societal divisions, they asserted that transformative mediation still had a grander vision. *Id.; see also* Isabelle R. Gunning, *Know Justice, Know Peace: Further Reflections on Justice, Equality and Impartiality in Settlement Oriented and Transformative Mediations,* 5 CARDOZO J. CONFLICT RESOL. 87, 87-88 (2004).

128. BUSH & FOLGER, *supra* note 10, at 46.

129. Judith Saul & Scott Sears, *A Relational Perspective on Multi-Party Practice, in* TRANSFORMATIVE MEDIATION: A SOURCEBOOK 405 (Joseph P. Folger et al. eds., 2010).

they include diverse views and strive to keep the process transparent,[130] and they define success as giving clarity and voice to the participants.[131]

Although Bush and Folger discourage combining their approach with others,[132] their work has inspired variations on the transformative model. For example, family therapist Richard Chasin and colleagues in the Public Conversations Project in Watertown, Massachusetts, have used the transformative theory but somewhat different practices to address policy conflicts over abortion, land use, and development.[133] They observe, "In polarized controversies, people are inclined to voice only that part of themselves that is consonant with the rhetoric associated with one of the poles. Conversations between opponents become confined to old positional arguments, attacks, defenses, and counterattacks."[134] To counter this tendency, Chasin and colleagues select the participants, pre-interview them, and set ground rules aimed at eliciting comments that are "authentic but not attacking."[135] They also use carefully prepared opening questions to urge participants to talk personally at the start, thus disrupting positions and stereotypes.[136] These practices exert more control over the conversation than a transformative mediator would, but the Public Conversations Project still salutes "Bush and Folger's broad achievement," a focus on improving the quality of the parties' relationship rather than on achieving settlement.[137]

---

130. *Id.* at 408.

131. *Id.* at 415.

132. Bush & Folger, *supra* note 10, at 228 ("Our experience is that combining models is not possible, because of the incompatible objectives of different models and the conflicting practices that flow from these diverse objectives.").

133. Chasin et al., *supra* note 13, at 323-25.

134. *Id.* at 334-35.

135. *Id.* at 338. "Feelings conveyed through character attacks, assignment of malicious motives, implied threats, insults to family, religion, race, nationality, and such may be so wounding as to offset any value to be gained by bringing those feelings to the table." *Id.*

136. *Id.* at 335.

137. *Id.* at 340 ("Nothing could be closer to our hearts than Bush and Folger's broad achievement: a conscientiously wrought and passionately expressed appeal to give priority to the

The University of Colorado Information Consortium has developed another transformative approach for use in large-scale policy conflicts.[138] It is called "constructive confrontation,"[139] and it aligns itself with the transformative model because it, too, aims at empowerment and recognition, although by different names.[140] It also rejects problem solving in favor of encouraging incremental changes in the parties, themselves.[141]

> Mediators (from Bush and Folger's point of view) or conflict specialists in general (from our point of view) would do better when they confronted difficult or intractable conflicts if they would abandon the search for resolution and pursue empowerment and recognition (in Bush and Folger's words) or constructive confrontation (in our words) instead.[142]

Like the Public Conversations approach to public conflicts, constructive confrontation's technique is more structured than Bush and Folger's.[143] Heidi and Guy Burgess have described the process in medical terms: diagnose a conflict, explore different treatment options, and monitor progress.[144] However, "[n]either constructive confrontation nor transformative mediation has a predetermined structure or step-by-step process beyond the most basic framework. Rather, both have a list of things to look for and respond to,"[145] and both look for incremental change.[146]

---

quality of the relational experience, and to the enhancement of the relational skills required when dealing with human beings at impasse. We agree strongly with this priority . . . . Most important, we must detach ourselves from the goal of settlement.").

138.   Heidi Burgess & Guy Burgess, *Constructive Confrontation: A Transformative Approach to Intractable Conflicts*, 12 MEDIATION Q. 305, 305 (1996). The project was formerly called the Conflict Research Consortium. *Conflict Information Consortium*, UNIVERSITY OF COLORADO AT BOULDER, http://conflict.colorado.edu/ (last visited Feb. 6, 2015).

139.   Burgess & Burgess, *supra* note 138, at 305.

140.   *Id.* at 307.

141.   *Id.* at 308.

142.   *Id.* at 307.

143.   *Id.* at 321.

144.   *Id.* at 307-09.

145.   Burgess & Burgess, *supra* note 138, at 308.

146.   *Id.* at 319-20.

In some other variations on the transformative model, mediators bring fairness and justice values into the process. For example, Isabelle Gunning, while expressing her attraction to the transformative model, urges mediators to actively address power imbalances between parties by asking parties to define "equality" and "justice"[147] and by later checking in to remind the parties of these definitions.[148] Ellen Waldman would proceed differently in cases where disputes involve public resources and public values.[149] She advises transformative mediators to advocate that settlements be guided by societal norms, rather than by norms generated by parties.[150]

These examples illustrate ways in which transformative mediation has been adapted to new situations.   Each focuses on empowerment and recognition as the overarching goals and understands those goals to be mutually reinforcing. Some adaptations, in addressing broader public issues, increase their control over selected aspects of the process. Still, none of the adaptations above contend with the overhearing audience that is the norm for a news interview.   The next section will take up the problem—and the potential—offered by the presence of the audience.

V. NEWS INTERVIEWS THROUGH A TRANSFORMATIVE LENS

One similarity between transformative mediators and news interviewers is their comfort with conflict. Describing the attributes of mediators in the transformative approach, Folger writes:

> Mediators are able to be "in the room" with escalating conflict and do not contain parties' conflict interaction by encouraging parties to avoid conflict, save face, offer forgiveness,

---

147.   Isabelle R. Gunning, *Know Justice, Know Peace: Further Reflections on Justice, Equality, and Impartiality in Settlement Oriented and Transformative Mediations*, 5 CARDOZO J. CONFLICT RESOL. 87, 90-91 (2004).

148.   *Id.* at 95.

149.   *See* Ellen A. Waldman, *Identifying the Role of Social Norms in Mediation: A Multiple Model Approach*, 48 HASTINGS L.J. 703, 703 (1997).

150.   *Id.* at 742.

109

or move to common ground. Instead, the mediator is comfortable allowing the parties to
explore the dimensions of difficult and divisive issues however they want to address
them, even if this means that the parties question or end their relationship, fail to reach an
agreement, or decide to escalate their conflict by pursuing it through an adversarial
process outside of the mediation.[151]

The same description would apply to news reporters, who are often "in
the room" with conflict and are professionally oriented not to influence its
outcome.

Like the transformative mediator, the reporter takes care to preserve
words and repeat them. In news interviews, the repetition of a party's
statements normally follows, not immediately, but soon after the words are
uttered, when the news organization quotes and publishes them. At that
point, the party does—or at least is able to—see or hear the words again by
reading or viewing the report.

In broadcast news, the technique of reflecting a news source's
statements also occurs in other ways. Once a party makes an important
statement, a reporter may recall it for the same party in a subsequent
interview. A brief look at transcripts of televised news interviews during
February and March 2014, as protests arose in Ukraine and as Russia
annexed Crimea, easily located examples. In one example, CBS News
correspondent Bob Schieffer reflected an earlier comment of Senator John
McCain, preserving and not muting it:

> SCHIEFFER: Senator, I want to ask you about something you said the other day. You
> said that President Obama was—and I believe these are your words—"the most naive
> president in history." Did you mean that literally, or how did you mean that?
> McCAIN: I meant it in my—in my time in public life. When you look at the so-called
> Geneva farce that was just a terrible joke, where we expected for Bashar Assad to come

---

151.   Folger, *supra* note 97, at 844.

to Geneva and arrange for his own transition from power when he was winning on the ground was ludicrous . . . . [152]

As in transformative mediation, the reflection let the interviewee consider and, if needed, restate or refine, what was said before. In transformative mediation, this kind of reflection would be used in order to help the speaker make an empowerment shift—establishing his presence on the record, letting him hear himself, and letting him modify his comments if desired.[153] In fact, in a news interview, further consideration is not merely allowed, but often requested.

Besides reflecting comments back to the speaker, reporters commonly confront an interviewee with a statement from an opposing party, especially if the statement is strong or provocative—the kind of statement that a transformative mediator might flag as an opportunity for reflection.[154]

> **DAVID GREGORY**: Congressman, let me ask about [former White House counsel] Kathy Ruemmler who is here. She worked for the president, tries not to get in all the partisan fights, but as [she] has said, "Look, this potential suit that the House speaker is pursuing has no standing, should be seen as beyond frivolous." What do you think?
> **REP. SEAN DUFFY**: Well, first of all, let's look at what the president's doing. He's taking historic action, never been done before, where he's waiving and suspending laws.[155]

In a transformative mediation, the reporter's restatement of an opponent's view would give the listening party a chance to re-hear the opposing view, possibly with less guardedness and more understanding.

---

152. Face the Nation Transcripts, CBS NEWS (Feb. 23, 2014), http://www.cbsnews.com/news/face-the-nation-transcripts-february-23-2014-mccain-jindal-omalley/.

153. *Institute for the Study of Conflict Transformation*, *supra* note 10, at 47.

154. *Id.* at 33 (listing samples of statements that present opportunities for empowerment and recognition shifts). These include statements that keep repeating the same point, trivialize the other's point, or assume the worst motives of the other. *Id.* It is worth noting, however, that transformative mediation would not carry messages between parties. *Id.* at 45.

155. *Meet the Press Transcript*, NBC NEWS (June 29, 2014), http://www.nbcnews.com/meet-the-press/meet-press-transcript-june-29-2014-n143826.

111

It is possible that the accurate repetition of a speaker's statements has the same effect in a news interview as in a transformative mediation. That is, the news source, seeing that he or she has been heard, may feel an incrementally increased sense of clarity, or empowerment. From this position, the source may reconsider and temper an extreme statement. Also, as in a transformative mediation, an opponent who now can see or hear the statement in a different voice, and at a safer remove, may be a bit more able to understand and empathize. Yet disputes are highlighted and summarized, and angry words are reflected in news media interviews every day. If these techniques, in a news context, were enough to prompt a disputant to experience more internal clarity and more empathy toward an opponent, then news interviews would be resolving conflicts routinely, and this is apparently not the case. In fact, in the examples here, the interviewees did not appear to rethink and temper their earlier comments nor did they demonstrate greater sympathy toward the views of others. Instead, they strengthened and defended their positions.

A.   The Audience Problem

The problem is that news interviews differ fundamentally from personal talk. Even though the broadcast medium requires the appearance of conversational ease, the news interview is highly choreographed[156] and geared to the ever-present but "unseen public."[157] It is *"the production of talk that is targeted for an overhearing audience."*[158] The interview is a structured question-answer talk, in which journalist and source take turns,[159]

---

156.   STEVEN CLAYMAN & JOHN HERITAGE, THE NEWS INTERVIEW: JOURNALISTS AND PUBLIC FIGURES ON THE AIR 28 (2002).
157.   MICHAEL SCHUDSON, THE POWER OF NEWS 75 (1995).
158.   CLAYMAN & HERITAGE, *supra* note 156, at 119.
159.   *Id.* at 95-119.

and the tone is generally civil, although questions may be adversarial or even hostile[160] and appear to have become more challenging in recent decades.[161]

Typically, news interviews engage representatives of institutions, most often political insiders;[162] therefore, the dialogue is goal-oriented.[163] "[M]ost live interviews are less a revelatory encounter than a kind of ceremonial ritual . . . a way for newsmakers to present themselves and deliver messages"[164] and for news organizations to excite audiences with spontaneous drama.[165]

In such circumstances, moments of empowerment and recognition, in the transformative sense, are likely to be rare. The founder of "peace journalism," Johan Galtung, acknowledged that "good conflict work is rarely done with millions watching and the parties playing to that enormous gallery."[166] Transformative mediator Judith Saul, who facilitates large group conversations on public issues in Ithaca, New York, points out that the presence of an audience can complicate dialogue by adding a layer of group

---

160. *Id.* at 188-237.

161. John Heritage & Stephen E. Clayman, *The Changing Tenor of Questioning Over Time*, 7 JOURNALISM PRAC. 481, 498-99 (2013).

162. CLAYMAN & HERITAGE, *supra* note 156, at 36.

163. Jeffrey Craig, *Dialogue and Dissemination in News Media Interviews*, 11 JOURNALISM 75, 79 (2010) ("News media interviews are a form of institutional dialogue, and, as such, distinguished against ordinary conversation or informal talk. Conversation is defined by a range of features (however problematic they may be) such as a greater sense of equal status between participants, equal rights to speak, greater degrees of reciprocity and assumptions about equal contributions to the conversation.").

164. KOVACH & ROSENSTIEL, *supra* note 38, at 131.

165. Ted Koppel, former anchor of ABC News Nightline, interviewed for the Television Academy Foundation's Archive of American Television, and recalled his role in pioneering live broadcast interviews with two or more opposing news sources, saying, "Roone [Arledge, then ABC News president] once described it as holding two wires or three wires together and watching the sparks fly." Videotape: Interview by Don Carleton with Ted Koppel, anchor and managing editor, Nightline, in Washington, D.C., at 18:03 in part four, Ted Koppel, (June 14, 2005) (on file with the Archive of American Television), *available at* http://www.emmytvlegends.org/interviews/people/ted-koppel# (transcribed by author).

166. Galtung, *supra* note 80, at 33 n. 11.

113

dynamics to the internal and interpersonal dynamics already present.[167] In a hypothetical television interview with allies watching, a spokesperson who suddenly glimpses some value in an opponent's viewpoint will likely worry, "Am I going to have the courage to speak up . . . without being marginalized and seen as a traitor?"[168]

Even in that hypothetical scene, however, Folger points out that empowerment and recognition are occurring, albeit silently, in the moment of that sympathetic glimpse.[169] This incremental change matters, he says, even if it is not publicly acknowledged.[170] "[R]ecognition can happen in thought, even when it is not overtly expressed in words. Careful and intentional interviewing can give interviewees mental pause, even if people do not express shifts in their thinking aloud."[171]

The typical television interview, however, has been offered as a classic example of "destructive debate" rather than dialogue.[172] Television news interviews are distorted, "repetitive, entrenched and rhetorical"[173] events in which interviewees can repeat talking points advised by political marketers.[174] Such stubborn repetition is just the kind of talk that would signal weakness and self-absorption to a transformative mediator. That means that, if viewed through the lens of transformative mediation, what

---

167. Telephone interview with Judith Saul, transformative mediator, facilitator, and trainer, Institute for the Study of Conflict Transformation (June 17, 2014).

168. *Id.*

169. E-mail from Joseph P. Folger, Professor, College of Education, Temple University, to author (July 24, 2014) (on file with author).

170. *Id.*

171. *Id.*

172. Chasin et al., *supra* note 13, at 326.

173. *Id.* at 325.

174. KOVACH & ROSENSTIEL, *supra* note 38, at 89-90 ("Talking points are preconfigured phrases and political-marketing buzzwords developed by communication specialists to manipulate public perception . . . . In the assumption that continuous repetition may convert assertion into belief, if not quite fact, the use of talking points has become standard procedure in Washington . . . .").

audiences observe in a contentious news interview is the lowest point on the
downward spiral of hostility—the place where both sides keep repeating
themselves and appear not to hear each other.   In the one-way
communication of broadcast and cable transmission, the audience cannot
respond. It can only observe powerlessly—or change the channel.

B.   The Audience Potential

News has been defined as communication that provides a threat to the
tranquility of the news consumer.[175] That is, news is news because it is
frightening.[176] Sociologist David Altheide charges that the point of the
interview—particularly when conducted for cameras—has changed from
gathering information to having an emotional impact on the audience,
promoting a "discourse of fear."[177] If that is so, then in transformative
terms, journalists work in a field that, by its upsetting nature, would tend to
disempower audience members, and journalists conduct interviews in a way
that tends to make that worse.

It is in this unnerving context that audience members encounter cross-
cutting views from a world outside of their own everyday experience, "an
often untidy and uncertain social world"[178] exposing them to "the possible
mental crises that may arise when we do not know what is real, or what to
think or believe."[179] This internal experience fits what the transformative
model would call a "loss of personal strength and clarity"—the opposite of

---

175. ARNO, *supra* note 26, at 8.
176. Gadi Wolfsfeld et al., *News about the Other in Jordan and Israel: Does Peace Make a Difference?*, 19 POL. COMM. 189, 191 (2002).
177. DAVID L. ALTHEIDE, CREATING FEAR: NEWS AND THE CONSTRUCTION OF CRISIS 101 (2002).
178. Stephen F. Ostertag, *Processing Culture: Cognition, Ontology, and the News Media*, 25 SOC. F. 824, 824 (2010).
179. *Id.* at 826.

115

empowerment.[180]   Transformative   mediation   would   expect   this
disempowered position to limit the ability of audience members to form
clear points of view, and it would expect them to respond by becoming
increasingly impervious to opposing views.[181]

Empirical studies lend support to this interpretation. Ambivalence is
uncomfortable,[182] and in experiments, people tend to respond by trying to
lessen the inconsistency in their own thinking.[183]   They look for new
information that will bolster their prior leanings and make them feel
better.[184]   They tend to avoid thinking about conflicting information.[185]
When news audience members do encounter a message that runs counter to
their own attitudes, they often resist it or mentally rehearse their reasons
against it.[186] To maintain their internal sense of order and security[187] they
may protect themselves by using a variety of mechanisms, including

180. BUSH & FOLGER, *supra* note 10, at 49-50.

181. This analysis has support from Gadi Wolfsfeld et al., *supra* note 176, at 191 ("News
stories about enemies are based on a common set of cultural assumptions, beliefs, myths, and
symbols all of which intensify the level of hostility and suspicion toward the other side. Journalists
routinely feed into these fears because they resonate with their audience.").

182. Jason K. Clark et al., *Attitudinal Ambivalence and Message-Based Persuasion: Motivated
Processing of Proattitudinal Information and Avoidance of Counterattitudinal Information*, 34
PERSONALITY & SOC. PSYCHOL. BULL. 565, 566 (2008).

183. *Id.* at 574 ("[A]mbivalent people are likely to process proattitudinal information at high
levels but are unlikely to process counterattitudinal information.").

184. Vanessa Sawicki et al., *Feeling Conflicted and Seeking Information: When Ambivalence
Enhances and Diminishes Selective Exposure to Attitude-Consistent Information*, 39 PERSONALITY
& SOC. PSYCHOL. BULL. 735, 742 (2013) ("Across two selective exposure studies, weaker,
conflicted attitudes yielded strong preferences for proattitudinal information when people lacked
knowledge about the issue. . . .   Presumably, this seeking of unfamiliar attitude-consistent
information was a means to resolve the tension of the ambivalent state (i.e. to bolster the strength of
the attitude).").

185. Clark et al., *supra* note 182, at 574 ("[I]t seems clear from these studies that ambivalent
people are likely to process proattitudinal information at high levels but are unlikely to process
counterattitudinal information.").

186. Prior, *supra* note 5, at 120.

187. Ostertag, *supra* note 178, at 826.

116

discounting the validity of the report.[188]   Hearing more news does not seem
to promote more receptivity.   Instead, studies find that increased news
coverage of a political topic increases the level of bias among audience
members, prompting them to become still more inclined to favor the facts
that support their side and resist those that do not[189] in a downward spiral
familiar to transformative mediation.

Fair and factual newscasts alone do not appear to be a remedy.   With
political leaders more ideologically split than ever,[190] the point-counterpoint
interviews that are common on major news broadcasts[191] confront audiences
with stark political choices.   Increasingly, ordinary Americans in the middle
feel disconnected from the forum.[192]   Many change the channel,[193]
abandoning the news in a continuation of the exodus that alarmed the early
advocates of public journalism in the late 1980s.   James K. Batten, then-
editor of the Knight-Ridder newspaper chain, characterized their flight using
language that evokes the self absorption described by transformative
mediation.   Batten called it "an inclination to withdraw into narrow, personal

188. *Id.* at 841-45 (listing the audience members' self-protective mechanisms as follows:
engaging news outlets selectively, imagining that they were better informed than the hypothetical
"everyday American" of their own cognitive construction, and using lay theories, "informal,
common-sense explanations," that journalism can be discounted because it aims primarily at profit,
sacrifices accuracy to audience demands, is influenced by the government, is limited by time
constraints, or is distorted by bias).

189. Jerit & Barabas, *supra* note 46, at 682 ("[W]hen a topic has implications for one of the two
parties, increasing the level of media coverage heightens partisan perceptual bias. In these instances,
there is an even stronger proclivity for learning politically congenial facts and for resisting
uncongenial ones. Indeed, our results suggest that extraordinary levels of media coverage may be
required for partisans to incorporate information that runs contrary to their political views.").

190. PEW RESEARCH CENTER, *supra* note 4, at 27.

191. Prior, *supra* note 5, at 103 ("Evening newscasts on the broadcast networks, long the most
widely followed news source, are mostly centrist with possibly a minor tilt in the liberal direction.").

192. *Id.* at 123.

193. *Id.* at 107 ("The culprit turns out to be not Fox News but ESPN, HBO, and other early
cable channels that lured moderates away from the news—and away from the polls.").

concerns and behave with indifference to our neighbors today and our communities tomorrow."[194]

The disappearing audience may deserve more sympathetic treatment, however. Political Science and Communication Professor Diana Mutz credits these apparently disinterested citizens with having tolerance and open-mindedness,"[195] being willing to change their minds on issues,[196] and being willing to downgrade the importance of their political identity for the sake of social harmony.[197] What may be lost when they avoid the public forum is their flexibility and their nuanced sense of things. "The voice of moderation is seldom very loud."[198]

Can journalism—rather than frightening or disorienting audience members—increase their capacity to clarify their own views while still recognizing the potential virtues of other perspectives? That is, in the transformative sense, can news interviews empower the public that journalists mean to serve? What would such interviews look like? The next section draws on transformative mediation principles to suggest some ideas and offer examples from the comparatively rare times when such interviews occur.

---

194.   ROSEN, *supra* note 48, at 23.

195.   MUTZ, *supra* note 12, at 125.

196.   Diana C. Mutz, *The Great Divide: Campaign Media in the American Mind*, 141 DAEDALUS 83, 85 (2012) ("Upon reflection, this pattern is not all that surprising; those most heavily interested and involved in politics in this country are also heavily partisan, highly committed to their choices, and thus unlikely to be dissuaded, regardless of any media to which they are exposed.").

197.   *Id.*

198.   *Id.* at 127.

VI. TRANSFORMING INTERVIEWS

Interviewing is possibly the least studied and taught skill in journalism.[199] Training in broadcast interviewing is typically brief and pragmatic, urging practitioners to prepare as well as possible within time constraints, to keep questions short, and to pay attention to technical details such as distracting background noise.[200] The field has wide room to further develop the study of interviewing and to add techniques that foster a functioning public forum. As a start, it could expand on the work of those relatively few news broadcasters who are creating safe spaces for audiences to consider other viewpoints. It could also offer incentives for more work in this direction.

A.   Safe Spaces for Differing Views

In the transformative mediation model, third-party neutrals begin by setting a context for the parties' own efforts at deliberation, decision making, communication, and perspective taking.[201] This is in line with Deutsch's ideal of constructive dialogue, which needs skillful facilitators, respected on both sides, who create an environment where opposing parties feel safe from

---

199. Bob Steele, *Interviewing: the Ignored Skill*, POYNTER (June 14, 2003), http://www.poynter.org/latest-news/everyday-ethics/talk-about-ethics/12413/interviewing-the-ignored-skill/.

200. *See, e.g.*, Casey Frechette, *What Journalists need to Know About Interviewing for Video*, POYNTER (Feb. 6, 2013), http://www.poynter.org/how-tos/digital-strategies/202713/what-journalists-need-to-know-about-interviewing-for-video/; *see also* AV WESTIN, BEST PRACTICES FOR TELEVISION JOURNALISTS (2000) (offering a chapter on interviewing that occupies not quite three of the handbook's ninety-seven pages and focuses primarily on interviewing the bereaved and attending to technical concerns such as camera use and editing).

201.   Folger & Bush, *supra* note 104, at 267.

harm and humiliation.[202] Where more parties are involved, the importance
of process design increases.[203]

Journalists can design a context for deliberation by establishing some
consistent programs that serve as safe spaces for considering new
viewpoints, and they can signal their intentions clearly to audiences.[204] Such
places would let audiences observe the high point of the communication
spiral, where people do hear each other and statements are not entrenched
and repetitive—where people have points of view but also recognize each
other's observations and maintain a relationship. This kind of conversation
might even respond to what Mutz has characterized as a current need for
"instruction, and specific norms, for how political differences should be
handled respectfully in informal discourse."[205] The following suggestions
are drawn from transformative mediation and other relational approaches to
conflict resolution.

1. Post a Sign

Audience members tend to look for news programs they feel they can
trust.[206] A few news broadcasts have demonstrated how to create and label
program segments as friendly, casual places for encountering unfamiliar
ideas. *Barbershop* and *Beauty Shop* were part of National Public Radio's
Tell Me More, a news program that aspired to present constructive dialogue

---

202. Morton Deutsch, *Mediation and Difficult Conflicts*, *in* THE BLACKWELL HANDBOOK OF
MEDIATION 361-62 (Margaret Herrman ed., 2006).

203. Saul & Sears, *supra* note 129, at 402.

204. Prior, *supra* note 5, at 120 ("News stories are experience goods, as users discover their
properties only upon exposure. This feature encourages news producers to create brands which
communicate to potential viewers what to expect.").

205. MUTZ, *supra* note 12, at 150.

206. Yariv Tsfati & Joseph N. Cappella, *Do People Watch What They Do Not Trust?:
Exploring the Association Between News Media Skepticism and Exposure*, 30 COMM. RES. 504, 518
(2003).

among diverse views.[207] On each segment, a regular group of commentators engaged in a sociable exchange of views on news, sports, and gossip[208] aimed at broadening and changing the public conversation.[209]

For example, in a July 2014 segment, host Jimi Izrael invited conservative commentator Lenny McAllister to join several regular participants, including international human rights lawyer Arsalan Iftihar and a contributing editor for the online publication *The Root*, Corey Dade.[210] The segment began with a sound bite from MSNBC, in which Texas Representative Henry Cuellar criticized President Obama for socializing in Denver but not going to the Texas-Mexico border, where thousands of unaccompanied children and teenagers had been entering the United States without legal permission.[211]

> IZRAEL: Lenny [McCallister], some folks are calling this Obama's Katrina moment . . . . Are you one of those folks? I mean, not to put you on the spot or anything, but . . .
> (LAUGHTER)
> McCALLISTER: Well, that's the whole purpose of having me on the *Barbershop*. . . . This is not Katrina . . . .
> IFTIHAR: This certainly is not Katrina. [T]his might be Republicans . . . trying to goad him into going, only to be used for potential political campaign ads.
> McCALLISTER: No it's not . . . .
> DADE: He's deported a record number of people . . . .

---

207. *About 'Tell Me More,'* NPR, http://www.npr.org/templates/story/story.php?storyId=5046 (last visited Jan. 26, 2015).

208. Douglas Hopper, *The Barbershop Revealed*, NPR (May 5, 2008), http://www.npr.org/blogs/tellmemore/2008/05/meet_the_guys_in_the_barbersho.html.

209. *Id.* (click on videotape for comment by Jimi Izrael at 0:08) ("I'm hoping that . . . what we've done here with the *Barbershop* is change . . . the language of how people expect politics to be discussed and change the type of people they expect to see discussing politics.") (transcribed by author).

210. *Tell Me More: Avoiding the Border: Is This Obama's Hurricane Katrina?*, NPR (July 11, 2014), http://www.npr.org/2014/07/11/330683502/avoiding-the-border-is-this-obamas-hurricane-katrina.

211. *Id.*

McCALLISTER: In 2012, he also allowed people to stay here in order to turn the tide
with the election. So there are two sides to that coin .... And we continuously want to
look at one side or the other. And again to Arsalan's point, at some point in time you're
going to get it either way from Republicans. So you have to make a choice. Do you
want to do things as a politician, or things as a leader?[212]

It is worth noting that, although from different political viewpoints, the
interviewees quickly discarded the obvious potential talking points—
whether the President had experienced his "Katrina" moment and whether
he was too pro- or too anti-undocumented immigrants. Instead, the question
raised was whether the President could rise above politics.

Barbershop opened each segment with friendly exchanges and moved
easily from politics to sports and other talk. The segment apparently
promoted recognition—in the transformative sense of that word—for at least
some listeners. One wrote in to an NPR website,

As a 62-year-old white woman in the Deep South, how else could I ever be a fly on the
wall in the barbershop? I LOVE those guys, even when I don't agree with a particular
statement .... [T]his group of highly intelligent, articulate, multifacted [sic] people has
helped me examine my own thoughts in a different light.[213]

Another effort at non-threatening political talk was presented by Al
Jazeera America, which described its program, *The Café* as "your chance to
listen in and contribute to a lively exchange of ideas—talking—not shouting
heads, from around the world."[214] Before the 2012 election, moderator
Mehdi Hasan hosted six American commentators with a range of views and
heard a point of agreement between Bruce Fein, who was an advisor to

---

212. *Id.*
213. *The End of Tell Me More*, ON THE MEDIA (June 20, 2014),
http://www.onthemedia.org/story/end-tell-me-more/ (follow "Show 36 comments" hyperlink). *On the Media* is an NPR media analysis broadcast produced by WNYC. ON THE MEDIA,
http://www.onthemedia.org/about/ (last visited Feb. 1, 2015).
214. *About the Café*, AL JAZEERA (last modified July 9, 2012),
http://www.aljazeera.com/programmes/thecafe/2012/05/20125161148521726.html.

Republican presidential candidate Ron Paul, and Clarence Page, a Chicago Tribune columnist and a supporter of President Obama.[215]

> **FEIN:** When the *New York Times* published the piece showing there are assassination plots right in the Oval Office, there's no outside review, there's no due process, the people get targeted are not accused of any crime . . . . If that article was written about what Bush and Cheney did, the left would be falling.
> **HASAN:** That's a fair point.
> **PAGE:** That, that is a fair point.[216]

To be sure, other political talk shows have their moments of humor and agreement as well, but these occur as deviations from the tone of crisis that is signaled by such names as *Hardball*, *Crossfire*, *Countdown*, and *The Situation Room*.

2. Invite Nonpartisan Talk

News broadcasts tend to select interviewees with strong, opposing points of view. The result is that they amplify the voices of the people least likely to have what negotiators would call a zone of possible agreement. Nonpartisan audience members are more likely to find their views reflected by interviewees speaking in nonpartisan ways. For example, Krista Tippett, in her *Civil Conversations Project* on American Public Media, interviewed former Republican Senator Pete Domenici and Democratic economist Alice Rivlin in October 2012, before a live audience at the Brookings Institution.[217]

The two, who had worked together on budget compromise, talked about partisan gridlock in Congress over the federal deficit:

---

215. *The Café: U.S. Still #1?*, AL JAZEERA (Oct. 23, 2012), http://www.aljazeera.com/programmes/thecafe/2012/10/20121022152246208655.html.

216. *Id.*

217. *On Being: Civil Conversations Project: Political Bridge People*, CIVIL CONVERSATIONS PROJECT (Oct. 25, 2012), http://www.civilconversationsproject.org/transcript-political-bridge-people-20121025.

MS. RIVLIN: I think the essential thing for people to grasp is that we're on an
unsustainable track. The debt is rising faster than our economy can grow. And almost
anybody can figure out that's a bad thing. . . .
MR. DOMENICI: I don't think what the public has to undertake to help us get this done
is as tough as it's being made. We're using words that scare people. "Cuts to Medicare".
. . . In fact, within five, six, [ten] years no one will know the difference, the programs
will have been impacted that small amount . . . .
MS. RIVLIN: [L]et me reinforce Pete's point. Because right now in this political
campaign, both sides are trying to scare people into thinking if you vote for this other
guy, it's going to be a terrible time. And Medicare is a good example. Both sides are
trying to scare seniors that their Medicare will be destroyed. . . . We need to adjust it at
the margin, but nobody's going to destroy it.[218]

In addition to finding collegial opponents, journalists can follow the
example of Chasin and colleagues, who choose mediation participants
precisely because they do not fit into the archetypes or stereotypes. Instead,
they look for people "whose own unique experiences differ in some respect
from others on their side."[219]

3. Welcome Uncertainty

One of the hallmarks of transformative mediation is its comfort with
confusion.[220] In posing questions to interviewees, reporters typically feel
compelled, as public representatives,[221] to ask aggressive questions,[222] and
they would rather been seen as too tough on an interviewee than too soft.[223]
But combative questions can prompt defensive responses. To elicit honesty
and to create a safe space for uncertainty, reflecting the ambiguities

218. *Id.*
219. Chasin et al., *supra* note 13, at 326.
220. Folger & Bush, *supra* note 104, at 272.
221. Steven E. Clayman, *Tribune of the People: Maintaining the Legitimacy of Aggressive
Journalism*, 24 MEDIA, CULTURE & SOC. 197, 198 (2002).
222. *Id.* at 211.
223. PEOPLE & THE PRESS, NEWS MEDIA DIFFERS WITH PUBLIC AND LEADERS ON WATCHDOG
ISSUES 30 (1995), *available at* http://www.people-press.org/files/legacy-pdf/19950522.pdf.

experienced by the overhearing audience, journalists can learn and develop a different kind of question. For example, in another edition of the *Civil Conversations Project*, Tippett asked Christian ethicist David Gushee and reproductive rights activist Frances Kissling to talk about their own areas of ambivalence.[224]

> **MS. TIPPETT:** What is it in your own position that gives you trouble? What is it in the position of the other that you're attracted to?
> **DR. GUSHEE:** One of the things I'm attracted to and have really learned a lot from in dialogue with Frances and others in the pro-choice community is the sustained knowledgeable commitment to the well-being of women. . . . A concern I have about my own side, and I know we're not going to focus on legalities, but I will say this. What the main activists in the pro-life or anti-abortion community want is an overturn of *Roe vs. Wade*. I am not at all convinced that if that were to actually happen that they would like the world that they would see . . . .
> **MS. TIPPETT:** OK. So Frances, what is it in your own position that gives you trouble? What is it in the position of the other that you are attracted to?
> **MS. KISSLING:** I'm generally troubled by the one-value approach to the question. That the only value that needs to be considered in both moral decision making and in legality is what the woman wants . . . . [E]ven though I don't think fetuses have an absolute right to life . . . I think fetuses have value. And I don't think you can make the fetus invisible in the abortion decision. . . .
> What I like about the position of people who are very strongly opposed to legal abortion, is that side of that movement . . . does have what David [Gushee] calls a consistent ethic of life. . . . I think that notion that there is a holistic need for respecting life and life processes is very attractive. And I think the arguments that are made about wanting to expand our sense [of] who is part of our community is a very attractive argument.[225]

---

224. *Pro-Life, Pro-Choice, Pro-Dialogue*, CIVIL CONVERSATIONS PROJECT (July 25, 2013), http://www.civilconversationsproject.org/transcript-prolife-prochoice-prodialogue-20130725.

225. *Id.*

4. Communicate Faith in the Audience

Transformative mediation takes an optimistic view of parties' competence and motives, expecting good faith and decency.[226] People in conflict are "only temporarily disabled, weakened, defensive, or self-absorbed."[227] Journalists need not abandon reality or their habitual skepticism in order to also allow for the possibility of decency in interviewees or, certainly, in the audience. On WNYC, a public radio station[228] in New York City, host Brian Lehrer expressed his confidence in his listeners as he began a year-end call-in segment by asking listeners to share some lessons they had learned.[229]

> **LEHRER:** What did you change your mind on in 2013? You know we pride ourselves here, at the Brian Lehrer show . . . on being a place where changing your mind is okay, where you don't have to be so ideologically pure all the time in whatever direction, where we ask you simply to bring your open mind with you every day. . . .
> **NOAH:** I started the year off as an ardent Obama supporter . . . and while I still agree with a lot of Obama's policies in principle, I find him to be much more incompetent and not up to the job of being President . . . .
> **SUE:** Pope Francis made me change my mind. I went back to . . . confession for the first time in, like say, forty years perhaps.
> **LEHRER:** Wow. Forty years.
> **SUE:** Yeah.[230]

Lehrer often draws on his diverse audience, which enables him to bring unusual voices to the air. In April 2014, for example, an avalanche sent ice

---

226. Folger & Bush, *supra* note 104, at 269.
227. *Id.* at 270.
228. Interestingly, WNYC, founded in 1924 by the City of New York, was specifically named in a footnote to the Hutchins Commission report as an example of "what government may do in domestic communications if it regards private service as inadequate." COMMISSION ON FREEDOM OF THE PRESS, *supra* note 33, at 5-6.
229. *Brian Lehrer Show: Open Phones: What Did You Change Your Mind About?*, WNYC (Dec. 18, 2013), https://www.wnyc.org/radio/#/ondemand/366997, at 2:15 (transcribed by author).
230. *Id.*

and snow down a slope of Mount Everest, killing sixteen Sherpa guides.[231] As his interviewees and callers discussed the, grueling, high-risk work of the guides accompanying climbers in the Himalayas, the studio telephone rang with calls from two ethnic Sherpas living in Queens, New York, who added their firsthand knowledge.[232] Lehrer said later, "[My] goal for my relationship with the audience is to sort of define everybody as 'us.'"[233]

B.   Incentives for Journalists

Journalists take their professional cues from other journalists, reading each other's work and rewarding each other's skills. The field especially honors those who fulfill the traditional watchdog role. A journalist who produces an excellent investigative broadcast news report may receive such honors as the George Polk Award,[234] the IRE Award,[235] the Philip Meyer Award and the Al Neuharth Awards,[236] or an award in any of the special investigative categories of honors bestowed by the Scripps Howard

---

231. *Brian   Lehrer   Show: Trouble   on   Everest*,   WNYC   (April   29,   2014), http://www.wnyc.org/story/trouble-everest/, at 10:09 and 14:10.

232. *Id.*

233. Ben Max, *Brian Lehrer, Interviewee: Iconic Radio Host Discusses His Career as 25th WNYC   Anniversary   Approaches*,   GOTHAM   GAZETTE   (July   1,   2014), http://www.gothamgazette.com/index.php/government/5129-brian-lehrer-wnyc-interview-25-year-anniversary, at 2:46 in fifth segment (transcribed by author).

234. LIU GEORGE POLK AWARDS, http://www.liu.edu/polk (last visited Jan. 11, 2015). The award was established by Long Island University in honor of a CBS News correspondent killed while covering the Greek civil war. *Id.*

235. *About   the   IRE   Awards*,   IRE: INVESTIGATIVE   REPORTERS   AND   EDITORS, https://www.ire.org/awards/ire-awards/ (last visited Jan. 11, 2015). Award winners are chosen by the Investigative Reporters and Editors organization. *Id.*

236. *Philip   Meyer   Award*,   IRE: INVESTIGATIVE   REPORTERS   AND   EDITORS, https://www.ire.org/awards/philip-meyer-awards/ (last visited Jan. 11, 2015) (for investigative reporting that uses social science research methods). These awards are also granted by Investigative Reporters and Editors. *Id.*

Foundation;[237] the Radio, Television, Digital News Association; the Gannett Foundation; and the Shorenstein Center at the Harvard University Kennedy School.[238]

Producing a story that promotes empowerment and recognition—that supports greater clarity and greater compassion for others—appears to garner no major award, although the Peabody Awards in broadcast journalism consider this kind of characteristic among other criteria.[239] In 2007, Brian Lehrer's program won a Peabody Award as "radio that builds community rather than divides,"[240] and an edition of Krista Tippett's program, then called *Speaking of Faith*, was awarded for a thoughtful exploration of Islam and was credited with "expanding our understanding of ancient and contemporary Islamic religiosity."[241] In 2009, broadcast producer and host Diane Rehm received a Personal Peabody Award for "thoughtful, civil discourse" on her long-running NPR program.[242] Such commendations appear to be rare, however.

---

237. *2013 Winners*, SCRIPPS HOWARD FOUNDATION, http://www.shawards.org/winners2013.aspx (last visited Jan. 11, 2015).

238. *Investigative Reporting Prize*, HARVARD KENNEDY SCHOOL: SHORENSTEIN CENTER, http://shorensteincenter.org/prizes-lectures/goldsmith-awards-program/investigative-reporting-prize/ (last visited Jan. 11, 2015).

239. *Judging*, PEABODY AWARDS, http://www.peabodyawards.com/judging (last visited Jan. 11, 2015) ("[T]he Peabody judges must ask themselves: Does this story matter? Does it inform us as citizens? Does it help us empathize with one another?").

240. *The Brian Lehrer Show (WNYC Radio)*, PEABODY AWARDS, http://www.peabodyawards.com/award-profile/the-brian-lehrer-show (last visited Jan. 11, 2015) ("Talk radio these days is so overwhelmingly polarized—or polarizing—that 'The Brian Lehrer Show' can seem more like an artifact than an anomaly. But it's very much in the present, reuniting the estranged terms 'civil' and 'discourse' five mornings a week like no other show on the air.").

241. *Speaking of Faith: The Ecstatic Faith of Rumi (Public Radio Stations Nationwide)*, PEABODY AWARDS, http://www.peabodyawards.com/award-profile/speaking-of-faith-the-ecstatic-faith-of-rumi (last visited Jan. 27, 2015).

242. *Personal Award: Diane Rehm*, PEABODY AWARDS, http://www.peabodyawards.com/award-profile/personal-award-diane-rehm (last visited Jan. 11, 2015) (noting that her program began in 1995 and continues on the air).

To advance journalism's constructive role in the public forum, journalism schools and organizations should institute annual awards specifically to recognize interviews and programs that promote clarity and compassion—empowerment and recognition in transformative terms—in interviewees, audiences, or both.   Such awards would encourage further development of interviewing theories and techniques.

Broadcasters Martin and Tippett are conscious of their role in providing models for audience members and other broadcasters.[243] They discussed the challenge in their own broadcast interview, when Martin was a guest on Tippett's program.[244]

> **TIPPETT:** The facts don't tell us the whole truth.
> **MARTIN:** I think . . . what you and I do is, in part, we offer a way for people to talk to each other that they might not have in their own lives. . . . I'm hoping what they will do is say to themselves, "I can do that, too." . . . This is not rocket science what we're doing here. . . . All we're really doing is setting an example.
> **TIPPETT:** We have so few templates about how you discuss a difficult issue. . . . But you can start it with a different framing question, and you can have a completely different conversation.
> **MARTIN:** The relevant aspect of [journalism] is the question and the listening, and anyone can do that.[245]

VII. AREAS FOR FURTHER EXPLORATION

This article suggests at least three areas for further research. First, adding non-combative, audience-empowering interviews to news programming may have revenue implications for news organizations. These are beyond the scope of this paper but are of considerable concern to the organizations. Some of the interview examples offered in this article reflect

---

243. *In   the   Room   with   Michel   Martin,* ON   BEING   (Oct.   2,   2014), http://onbeing.org/program/feature/in-the-room-with-michel-martin/6799.

244. *Id.*

245. *Id.* (click on audio at 31:20).

the *Washington Post*'s media critic, who called it "forced and forgettable."[252] *Crossfire* fell to network budget cuts in October 2014.[253] There is ample room for scholars and practitioners to explore not only how to create successful, constructive news dialogues but also how to pay for them.

Second, this article has focused on interviews, the format in which news organizations provide the least context for audience members. More often, television news is presented in the form of "packages," and radio news in "wraps," in which reporters introduce topics, select and summarize facts, and write passages that lead audience members through the various brief "sound bites" of sources. Likewise, in written news accounts in print or electronic media, the journalist contextualizes the raw material. Journalists write stories. The way that journalism seeks, gathers, crafts, and amplifies the stories that a society tells itself has implications for conflict and compromise in the public forum. Journalistic storytelling deserves further investigation in this light.

Finally, as journalism rethinks and redefines its business model and itself,[254] researchers should undertake interdisciplinary projects to better understand what kind of news coverage best serves the public forum and how to achieve it in a time of rapid change in information and communication technology. The question is at the intersection of conflict resolution, journalism, communication studies, political science, social psychology, and doubtless other fields as well. Journalists and their audiences can use the help.

---

252. Erik Wemple, *'Crossfire's' 'Ceasefire' Segment: So Lame!*, WASH. POST, Sept. 13, 2013, http://www.washingtonpost.com/blogs/erik-wemple/wp/2013/09/13/crossfires-ceasefire-segment-so-lame/.

253. Brian Steinberg, *CNN Cancels 'Crossfire,' Other Shows in Wake of Turner Cuts*, VARIETY (Oct. 15, 2014), http://www.variety.com/2014/tv/news/cnn-shutters-select-shows-in-wake-of-turner-cost-cuts-1201331097/.

254. *See, e.g.*, Donica Mensing, *Rethinking [Again] the Future of Journalism Education*, 11 JOURNALISM STUD. 511, 512 (2010) (recommending a re-examination of journalism practices and a new focus on community needs and relationships).

VIII. CONCLUSION

Journalism has the potential to introduce news audiences to alien viewpoints in a way that supports a functioning public forum, where people have clear ideas but can still hear and consider the ideas of others. Transformative mediation offers the news media a theory of conflict and a set of practices that can be adapted toward this end without violating the norms and ethics of journalism. The news media can, among other things, create regular interview segments where interviewees can acknowledge areas of agreement, and activists can reveal their own uncertainties. Journalism schools and organizations can reward such projects, and practitioners and academics can join across disciplines in undertaking research into the kind of talk that best promotes a good public forum.

The recommendations offered here are not meant to supplant the important role of the news media in gathering and verifying facts, but rather to supplement that role in ways that make those facts more useful and meaningful to a broader audience. The recommendations are consistent with some of the reforms of civic journalism and peace journalism. Still, they avoid the kinds of practices that have stirred professional and ethical objections to those movements. For example, these recommendations do not involve journalists in such activities as promoting peace settlements, organizing town meetings, or taking sides in a violent conflict. The recommendations do reflect a new thought about the role of journalism in interviewing opposing parties, and they suggest ways to fulfill that role: developing programming that supports audience members in moving toward greater personal clarity and empathy—toward empowerment and recognition—so they can better decide what to do next.

POLITICS | Articles & More

What Are the Solutions to Political Polarization?

Social psychology reveals what creates conflict among groups and how they can come together.

BY LEE DE-WIT, CAMERON BRICK, SANDER VAN DER LINDEN | JULY 2, 2019

What drives political polarization?

Is it simply disagreement over the great issues of the day? Not necessarily. Recent research by the More in Common Foundation found that more than three-fourths of Americans support both stricter gun laws and a pathway to citizenship for undocumented immigrants brought here as children. Roughly the same number of Americans agree "that our differences are not so great that we cannot come together."

Are they right?

The **Greater Good Toolkit** includes 30 science-based practices for a meaningful life.
Here is what a recent customer shared:

★ ★ ★ ★ ★

"This "toolkit" really is outstanding. It gives practical suggestions of practices and gives all the evidence-based research behind the practice. Words and ideas can be inspiring but providing easy access to the science behind the specific ideas is extremely helpful. Plus is it visually lovely."
- Michele R.

THE GREATER GOOD TOOLKIT →

From the GGSC to your bookshelf: 30 science-backed tools for well-being.

The More in Common results could be interpreted to suggest that we can build bipartisan support for specific policies by focusing more on their boring nuts and bolts. Unfortunately, however, voters don't evaluate policies in isolation. Research has highlighted that people actively use partisan cues when evaluating different policies.

For example, a study by Carlee Beth Hawkins and Brian Nosek shows that labeling policies as "Democrat" or "Republican" can influence policy support, depending on the implicit bias of participants toward each party. A 2017 study by David Tannenbaum and colleagues finds that support for policy "nudges"—such as changing 401k retirement accounts to opt-out rather than opt-in—was heavily influenced by whether they were framed as supporting the goals of the Democratic or Republican party. This was true of regular U.S. citizens and for senior government leaders. Similarly, a 2018 study by Leaf Van Boven and colleagues finds that the majority of Republicans agree that climate change is happening—but their support for policy solutions declined when presented by Democrats.

In other words, people like policies proposed by members of their own in-group—and they don't like ideas generated by out-groups. This dynamic is not new. Since the 1950s, social psychologists have tried to understand what pits groups against each other—and today, they're applying these insights to figure out what is happening in the United States. This research doesn't provide definitive answers, but it does suggest some potential solutions, from changes to the voting system to the development of common goals that might enable groups to work together.

How morality becomes partisan

The More in Common report illustrates that some of the most divisive topics often involve deep moral beliefs. For example, different political groups are very polarized on beliefs about responsibility, such as "people's outcomes in life are determined largely by forces outside of their control," or "people are largely responsible for their own outcomes in life." Similarly, liberals and conservatives are

very divided on questions of whether parenting should focus on cultivating a child's curiosity versus good manners, or independence versus respect for elders.

In a new study published this year, Annemarie S. Walter and David P. Redlawsk directly pitted people's moral concerns with their partisan identity. They presented 2,000 participants with examples of different moral violations by different actors. Based on previous research, Walter and Redlawsk had thought that the nature of the moral violation might be the most significant factor in people's evaluations, as there are reasons to think that liberals and conservatives are concerned with some moral violations more than others. What they found, however, is that it *wasn't* the nature of the moral violation that was most important. Instead, it was the political allegiance of the violator. Democrats in the study were prone to giving Democrats a pass; the same was even more true of Republicans.

Special Event: Bridging Differences

Join us to learn research-based strategies for
dialogue and understanding

This partisan influence on policy preferences and moral
judgements is a cause for both hope and concern. On one

Register Here

hand, it reiterates a point made by Daniel Yudkin in a
New York Times op-ed about the More in Common report:
that the U.S. may actually be less politically polarized
based on certain moral or policy issues—at least when
there aren't clear partisan associations. On the other
hand, it highlights that as soon as a moral or political
issue becomes associated with a particular party, it can
become polarizing.

This is why it increasingly feels like U.S. politics has
entered into a vicious cycle, whereby the moral and

emotional language used to galvanize one side is directly antagonizing the other. The us-and-them nature of the debate has led to such a breakdown of trust that even hearing a policy proposed by the other side can be enough to trigger opposition to that policy. New policies (whatever their merit) can therefore quickly become symbols of conflict for the two sides to rally around.

What are the solutions?

This suggests that while there might be various political seeds that have helped drive the recent spike in polarization, it has gotten to a point where polarization is being exacerbated by some of the psychological processes that shape how we interpret identity and groups. This is a significant point to understand because it highlights that if we are to address polarization, we need to think not just about political solutions, but also solutions that are grounded in our understanding of social psychology.

1. Intergroup contact. The "contact hypothesis" suggests that getting to know each other can reduce prejudice between groups. However, social contact can be done

well and done badly. As we discussed in a previous article, following political opponents on Twitter can make people more extreme in their political views. It turns out that many conditions have to be met for contact to reduce prejudice, including having contact be sustained, with more than one member of the group, including a genuine exchange of ideas, and between individuals of similar social rank. These conditions have been very difficult to meet in designing social policies.

One promising civic model for enabling more meaningful contact between groups in conflict involves "Citizens Assemblies," where representative citizens are brought together to deliberate over challenging social or political issues. These assemblies can be thought of as a kind of jury duty for political deliberation, and they offer a platform for different groups to discuss issues in a way that can highlight where common ground exists and how it can be acted upon.

For example, Ireland has run several Citizens Assembles since 2016 that made policy recommendations that have been credited with advances in Ireland's approach to

climate change. Indeed, participants in a recent Citizens Assembly on Brexit, run by Alan Renwick and colleagues at University College London, came to a compromise that could resolve the current impasse surrounding the U.K.'s decision to leave the European Union.

Shared Identity

Encourage generosity by finding commonalities between people

2. Perspective taking. Perhaps one of the most important aspects of contact is that it might enable one to see things from another's perspective. The promise of perspective taking was recently illustrated in an experiment to attempt to change support for issues faced by transgender minorities. In this intervention, a brief

exchange exploring a range of issues from the perspective of a trans individual was sufficient to shift people's attitudes on this controversial topic. Indeed, the attitude change seemed to persist even six months later, which is unusual for brief psychological interventions.

In his book *The Better Angels of Our Nature*, Pinker argues that the printing press may have had an important role in increasing levels of empathy following the Enlightenment by making it easier to read stories framed in the perspective of others. Indeed, Pinker speculates that some of the literature written from the perspective of black slaves may have been instrumental to the abolition of slavery.

Considering the revolution in communication technology in our lifetimes, social media may have done more to promote taking sides than seeing the world through the eyes of another. Social media companies, and the governments that regulate them, clearly need to reflect on the extent to which these platforms encourage "side taking" instead of "perspective taking."

3. Superordinate goals. One of the clearest solutions from the psychological literature is that identity-based conflicts require common goals or a "superordinate" sense of identity to bring people back together. In other words, we need a large sense of ourselves that is able to bridge smaller differences. This need to create a superordinate identity has clearly been intuitive to rulers for centuries, who would use various traditions and ceremonies to help build alliances between different countries and cultures.

Of course, superordinate goals also come with a potential risk. Whenever we form an in-group, we also create out-groups. As Richard Dawkins recently tweeted:

National pride has evil consequences. Prefer pride in humanity. German pride gave us Hitler, American pride gave us Trump, British pride gave us Brexit. If you must have pride, be proud that *Homo sapiens* could produce a Darwin, Shakespeare, Mandela, Einstein, Beethoven.

Unfortunately, drawing a parallel between Trump and Hitler is perhaps itself an illustration of the polarized nature of modern discourse. Dawkins does have a point, however: The use of a superordinate identity such as American or European has potential risks. So, should we just think of ourselves as humans—or is the idea of "humanity" too abstract? Former British prime minister Theresa May famously criticized such a universalist perspective, stating: "If you believe you are a citizen of the world, you are a citizen of nowhere."

Is that true? Research suggests that a universalist perspective might well have underappreciated benefits. Sam McFarland and colleagues recently reviewed this topic and found that those who identify highly as citizens of the world are indeed more empathic. Of course, those who are more empathic might simply identify more as international citizens. This idea warrants further testing, particularly as McFarland and colleagues identify several factors that might serve to further develop this sense of international citizenship.

4. Proportional voting. While searching for psychological solutions to polarization, it's important not to ignore the context in which political decisions are made, and to think about the way in which different political systems will engage with, and exacerbate, aspects of our psychology.

The U.S. is one of the few countries to be dominated by just two political parties. This fact is almost certainly a reflection of the "winner take all"/"first past the post" voting system. Many countries employ a proportional (or mixed) system, which means that if a party gets 5 percent

of the popular vote, they will receive 5 percent of the seats in a given representative body. In the U.S., this party would almost certainly get no representation—which could worsen the us-and-them dynamic of the U.S. political system.

SHARE YOUR BRIDGING DIFFERENCES STORY

The GGSC's **Bridging Differences** initiative aims to help address the urgent issue of political and cultural polarization. Do you work to help people or groups bridge their differences, whether as a mediator, organization leader, educator, politician, workplace manager, or otherwise? **Fill out this short survey** and let us know how we can help.

For now, however, there isn't systematic evidence comparing the extent of identity-based politics with the political system used in the U.S. Unfortunately, that kind of large-scale, cross-country research is

often the most difficult to obtain, but could be exactly what we need to understand how different voting systems might influence polarization.

That said, there is evidence that more proportional systems have higher levels of voter turnout (at least for supporters of smaller parties). In turn, that increase in turnout is correlated with citizens being more likely to report feeling that their vote makes a difference. This doesn't necessarily stop politics becoming less polarized, but it might make it harder for the extremes to come to dominate.

5. Voting for policies, not for parties. Another potential solution to identity-based policy preferences is to hold direct referendums on specific issues. Among large territories, California and Switzerland both regularly use referendums to address complex policy topics. Referendums are used less frequently in other countries to try to resolve controversial topics, as was the case with gay marriage in Australia or voting reform in New Zealand. When designed well, referendums might cut

across existing partisan divides, and if a clear majority is reached, they can signal a new social norm that can help a country move forward.

For those who are familiar with the fallout from the recent Brexit referendum in the U.K., however, this suggestion would probably seem a little laughable. Contemptuous, even. Indeed, there is evidence that the referendum in the U.K. has itself spilled over into a new form of emotional polarization, as recent data from YouGov highlights that (especially younger) "Remain" supporters would not want to see a close relative marry a "Leaver" (a member of the opposing political camp). There's another problem as well: While Australians did indeed vote to legalize same-sex marriage, it could have gone the other way; allowing the majority to vote on the civil and human rights of a minority is very risky.

Like many complex political systems, however, referendums can be designed well and designed badly. In countries with more established systems of direct democracy, the U.K. referendum wouldn't have even been legal. For example, in Switzerland, referendums have to

be about precisely defined changes to the law, not vaguely defined outcomes. In New Zealand's referendum on the voting system, an independent educational body was created to inform both sides of the debate without taking a position (as the British government controversially did during the Brexit referendum).

The psychological impact of more direct voting systems is worthy of further enquiry. When poorly implemented, referendums risk causing new fault lines along which polarization can manifest. When well implemented, referendums might cut across existing lines of polarization and help establish a new social norm that can move a country forward.

From reframing issues to tap into a superordinate sense of identity, to promoting forms of contact that encourage perspective taking, social psychology does offer some useful ideas for thinking about how to tackle polarization. Furthermore, social psychology provides insights into the potential implications of different kinds of voting systems and the way in which they might exacerbate or diminish identity-based politics. As we have been careful to try and

illustrate, however, experiments in social psychology do not yield off-the-shelf solutions that would be effective in all political contexts. Nevertheless, the farther that modern politics sinks into a self-fulfilling cycle of identity-based polarization, the more we'll need new insights from social science.

Get the science of a meaningful life delivered to your inbox.

Email Address

Sign

About the Authors

Lee de-Wit

Lee de-Wit, Ph.D., is a lecturer in political psychology at Cambridge University. He is also the author of *What's Your Bias?: The Surprising Science of Why We Vote the Way We Do*.

Follow

Cameron Brick

Cameron Brick, Ph.D., is a social psychologist who studies how people react to shared dilemmas such as political polarization or climate change using surveys and experiments. He is a research associate at the University of Cambridge, U.K., in both Maths and Psychology, and is a by-fellow at Churchill College.

Follow

Sander van der Linden

Sander van der Linden, Ph.D., is assistant professor of social psychology and director of the Cambridge Social Decision-Making Lab at the University of Cambridge. He is also a fellow of Churchill College, Cambridge.

Follow

By John M. Livingood

Reframing
and Its Uses

A close look at "reframing" — one of the most powerful tools in a conflict resolution practitioner's tool kit. The author explains the concept and explores the many opportunities for its effective use.

I n its broadest sense, "reframing"—a term of art in dispute resolution circles—is a realignment of "a frame of reference."[1] In negotiations and conflict resolution, it is a powerful tool that has many uses. In this article, the discussion of reframing is limited to the technique of restating or rephrasing statements and concepts in order to advance the goal of reaching an agreement and resolving conflicts and disputes. (This is not meant to discount the value of other forms of reframing, including the reframing of actions or behaviors.) Reframing can be used negatively to frustrate and impede settlements and the resolution of conflicts, but that will not be considered here.

All the participants in a mediation may engage in reframing. An advocate may reframe statements to further negotiations or channel them in a particular direction for the purpose of achieving the client's goals. A conflict resolution practitioner may, for example, reframe a statement by a party to the mediation in order to elicit agreement on some issues, or to bring out the underlying interests of the parties. The practitioner also may encourage the parties to reframe their own statements.

Generally, when a mediator reframes a statement, or encourages a party to do so, the technique is being used intentionally for a specific purpose. However, reframing also may occur inadvertently, for example during the course of back-and-forth bargaining or during a conversation with the mediator in a private caucus. Mediators are susceptible to inadvertent reframing when they attempt to paraphrase a statement with the sole intent of demonstrating active listening.

In the context of negotiations, William Ury has defined reframing as "redirecting the other side's attention away from positions and toward the task of identifying interests, inventing creative options, and discussing fair standards for the selection of options."[2] This definition, founded on interest-based bargaining principles, parallels interest-based approaches in mediation and conflict resolution generally. But reframing for this purpose is not an essential mediation or conflict resolution technique.

In addition to the uses of reframing suggested by his definition, Ury suggests using reframing tactics to address stone-walling, verbal attacks, and tricks in bargaining.[3] For example, he proposes reframing a personal attack into a question directed toward remedying the problem. Here is an example:

Original Statement: Don't you know any better than to submit a proposal that will never fly?

Reframed Statement: You may have a point there. How would you improve the proposal to make it fly?[4]
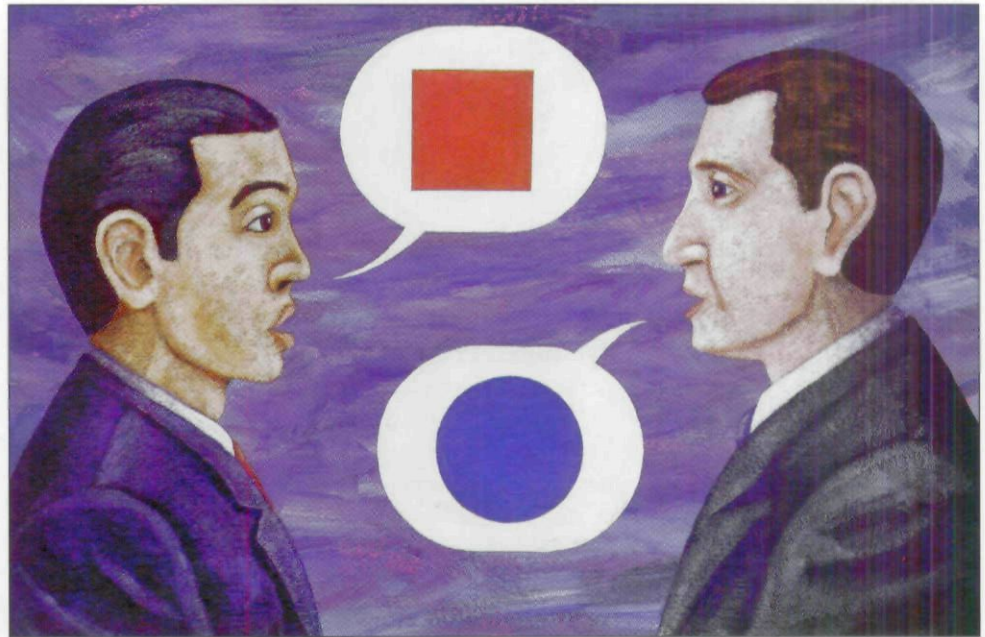
Reaction to Reframing

Certain sectors of the conflict resolution community have approached reframing very cautiously. While many dispute-resolution professionals stress the importance of including the interests and needs of the parties in reframed statements, others voice concern that mediators may intentionally or unintentionally impose their perspectives or attitudes on the parties. Notwithstanding these

John Livingood has been mediating a wide variety of disputes for more than six years. He has extensive training and experience in various styles of mediation. He is a mediator for the National Mediation Board. In addition, he has mediated for the U.S. Postal Service and the EEOC. The views expressed here are solely those of the author and should not be attributed to any other person or public or private entity.

concerns and consistent with the theme of encouraging the free flow of thought, conflict resolution practitioners generally have embraced reframing and some have refined and developed additional reframing techniques. One such technique, called "metaphoric reframing,"[5] attempts to find a new or altered metaphor to describe a situation or concept. It is an extremely difficult technique to master and it requires a great deal of sensitivity to the underlying meaning that the metaphor has for the parties.[6] Extreme caution should be exercised in using this very subtle technique.

There are at least six reasons to use verbal reframing. However, they are not mutually exclusive. In any situation there may be more than one reason to use reframing to achieve one or more particular effects. Reframing can be used to:

- achieve understanding and/or clarify a statement;
- help the author of the statement and/or the other participants achieve a different perspective;
- neutralize language;
- construct a joint or common goal, statement of the problem or issue in dispute;
- create a new relationship paradigm; and
- move the resolution process into a more focused phase.

The use of reframing for each of these purposes is discussed further below.

Understanding/Clarification

In negotiations and in conflict resolution processes, reframing is often used to help the parties and the dispute resolution professional gain a better understanding of the issues and the parties' positions, wants, interests and needs. The technique may conjure up the proverbial "peeling back the layers of the onion."

Reframing is particularly helpful in narrowing the issues in dispute and determining the true essence of the matter under review. For example, a mediator can reframe a statement made by a party that would clarify whether a particular issue is or is not in dispute. Or the mediator could ask the author of the original statement a question that would call for her to reframe it. Here is an example of question-induced reframing:

Original Statement: Our members are highly skilled and deserve to be paid for their skills.

Request to Reframe: Are you saying that all your members are highly skilled and all deserve higher pay?

Reframed Statement: Our mechanics and technicians are the ones I am talking about; they deserve higher rates.

Ury suggests asking problem-solving questions to focus attention on the interests of each side, the options for satisfying those interests, and the standards of fairness for resolving differences.[7] In Ury's view, "Why?," "Why not?," "What if?," and "What makes that fair?" are among the most useful questions. Additionally, he suggests asking the parties for their advice. For example, "Why do you want that? Help me understand the problem you are trying to solve."[8]

When the reframing technique is directed at discovering the parties' underlying interests, Christopher Moore, a renowned mediator and author, characterizes it as "testing."[9] In other words, the mediator listens to a party's statement and then states what he heard, and the interests the party was expressing. Gradually, through trial and error, the mediator helps identify the party's underlying needs and interests. In the following example, the mediator uses this questioning technique.

Original Statement: You must honor seniority when you award positions to employees.

Request to Reframe: Are you saying that I have to give a position to someone who is not qualified if he or she is senior?

Reframed Statement: Assuming the applicants are qualified, you must award the position to the most senior.

Through question-induced reframing, the parties may develop a new or varied perspective, which may give them greater insight into the issues. It can give rise to a new paradigm in thinking and to new opportunities for constructive engagement and resolution.

Creating New or Altered Perspective

Moore notes that disputing parties each have their own individual and subjective understanding of the issues in dispute and the basis of their conflict. Their perceptions are "images of reality," which are, in fact, simply interpretations of reality. A classic example of this phenomenon is the

Reframed Statement: Your sick leave policy requires a doctor's note on one-day illnesses. It's too costly to go to a doctor for a one day illness.

Dwight Golann, in *Mediating Legal Disputes*, describes how a mediator attempts to change the parties' perspective about a dispute and each other.[12] The mediator does this by giving each side information about the other's perspectives and arguments, questioning each side's views, and offering alternative perspectives on the situation.

Reframing to change a party's perspective may be influenced by philosophical approaches to conflict resolution. If the dispute resolution professional believes that disputes should be resolved based on the parties' real interests, then reframing will be used to orient the parties away from posi-

While the parties will inevitably have their own perspectives of the issue and possible solutions, reframing can contribute to the resolution process by focusing their efforts on the resolution of a common problem.

"glass is half empty; the glass is half full" view of the world.

While the number of potential interpretations is large, Moore observes that an individual's mindset generally allows only one "possible, reasonable, permitted view," which suggests only "one possible, reasonable permitted solution."[10] As a result, a party to a dispute will need help in seeing any other point of view or other possible solutions.

Using reframing, a conflict resolution professional can elicit statements from the parties without the rhetoric and imagery associated with party positions, and thereby assist the parties in hearing and understanding other points of view, perhaps for the first time. The author of the original statement may even begin to look at the problem in a different light.

Reframing that focuses on the language that the parties themselves use forces the parties to really hear their own statements as the mediator or the adversary heard them. As a result, the parties may choose different words, words that are less inflammatory, or more directly related to the issues and to their underlying interests, to explain themselves.[11] Here is an example:

Original Statement: Your sick leave policy is unfair.

tional statements toward expressing their needs or interests.[13] One commentator has noted that since reframing provides the mechanism to expand the issues, it can provide the parties with more bargaining power, since broader interests allow a greater spectrum of solutions.[14]

Reframing is problematic, however, when attempted in a dispute involving value-related issues, such as guilt or innocence, what norms should prevail in a social relationship, what facts should be considered valid, what beliefs are correct, and what principles should guide decision makers.[15] While partisanship and bias inevitably will pervade value-laden disputes, Moore suggests that reframing can still play a role. One is to transform a value dispute into an interest dispute. A second is to identify "superordinate" goals. This refers to goals with which all of the parties can identify and be joined in a cooperative effort.[16] Using superordinate goals is similar to finding "joint or common goals," discussed below.

Neutralization of Language

Perhaps the most familiar and frequent use of reframing is to reconstruct a statement to neutralize language by eliminating rhetoric, inflammatory language, partisanship and bias. The parties to a dispute tend to use language that is judgmental, positional, and biased toward their subjective view. A statement in need of reframing may be

obvious or subtle in its bias; it also may be personally insulting; or it could be impersonal yet value-laden. A conflict resolution professional who uses reframing will tend to translate the offending statement into "neutral terms to remove bias, positions and judgment."[17] Generally, by using neutral language, the mediator will be able to influence the parties to address the issues in a more neutral and positive manner.[18]

Bernard Mayer calls reframing with this purpose in mind "detoxification." He notes that the unproductive language is removed while the "essential interests" of what is expressed is retained.[19] Mayer asserts that the "hardest part of this type of reframing is making sure that neither the underlying concerns nor the intensity with which they are felt get minimized or discounted in the reframing process." He says that "the most common tactic is to replace value-laden language and positional demands with interest-based formulations."[20] Here are two examples of detoxification reframing in different contexts.

Example 1.

Original Statement: The company must pay for violating seniority rules.

Reframed Statement 1: If seniority was violated, there should be an appropriate remedy.

Reframed Statement 2: If seniority was violated, what is the remedy, if any?

Example 2.

Original Statement: Mr. X has lied to us and is untrustworthy.

Reframed Statement: Mr. X has given me information previously that has not been correct, and I am hesitant to rely on his current statements.

The original statement in Example 2 above is an extremely judgmental statement about Mr. X. The reframed statement eliminates the "liar" charge and replaces it with the author's perception of the facts (Mr. X has previously given incorrect information). The reframed statement also offers the author's feeling about Mr. X ("I am hesitant to rely on him"), removing the toxic "untrustworthiness" label in the original statement.

Mayer gives the following example in the context of a marital dispute.[21]

Wife's Statement: He could care less about our children. All he is worried about is how

> # When parties to a dispute use language that is judgmental, positional, and biased toward their subjective view, the mediator can translate that language into "neutral terms."

much of his precious money is going to have to pay support.

Mediator's Reframed Statement: You don't think that he is really motivated by your child's welfare, but you are clear that he wants to minimize how much money he has to pay.

Mayer also gives this example in the employment context.[22]

Employee's Statement: Hell will freeze over before I agree to work with that jerk again. It was torture last time we were on the same team, and I won't subject myself to his arrogance and sadism again.

Mediator's Reframed Statement: You had a very bad experience working together, and you do not want to repeat it. In particular you felt exposed to certain behaviors and attitudes that you do not feel you should have to deal with in the workplace (or elsewhere).

In the following example, a party is "apologizing" for a previously made statement.

Original Apology: I am sorry that your perception of what I said caused you any sense of harm or embarrassment.

Reframed Apology: I am sorry that anything I said caused you any sense of harm or embarrassment.

The original apology may not seem to be "toxic" on its face, but place it in the context of a sexual harassment case with the male supervisor making the initial statement to the female employee. The words "your perception" become a virtual Molotov cocktail; they not only defeat the attempt at an apology, they add fuel to the fire.

As the above examples show, reframing is very effective in neutralizing inflammatory language in statements that have been made. It can also be used in a preemptive fashion, to set forth the parameters of the dialogue between the parties, thereby avoiding judgment, bias or positional statements. For example, prior to formally embarking on their tasks, mediators usually will set ground rules that may include a warning

against using negative comments and personal attacks. Some mediators may instruct the parties prior to the actual mediation sessions in how to identify their interests, as distinguished from their positions. The goal is to encourage the parties to consciously reframe their thoughts prior to articulating them.

Identifying the Joint or Common Problem

As Moore has noted, each of the parties inevitably will have his or her own perspective on the issues and the possible solutions to a dispute. Verbal reframing can contribute to the resolution of a dispute by focusing the parties' efforts on the common problem and on identifying common goals. Moore also suggested this when he recommended using reframing to identify "larger superordinate goals" in value-related disputes, and when he discussed framing joint statements of the problem that incorporate the parties' individual and joint interests.

Problem-solving questions can focus attention on interests, potential options for settlement, and fairness standards.

Reframing in order to construct a statement of the common issues in dispute requires the interests of all the disputing parties to be included in the comprehensive statement. Once the parties agree on the statement, they can commit to working together on the common problem because they believe their needs will be respected, if not met, by the solutions that will be developed.[23]

Mayer employs the term "definitional reframing" to describe the process of redefining the issues or conflict so that a more integrative resolution can be found. This involves "conceptual reframing," which often takes the form of presenting an issue as a mutual problem to be mutually solved. This approach is similar to Moore's in that it incorporates the essential needs and concerns of all parties in the restatement of the common problem. Here is an example of definitional reframing in the context of a labor negotiation.

> *Original Statement*: We need a wage increase for our members to get this agreement ratified.
>
> *Reply*: We can not pay the wage increases you are looking for.
>
> *Reframed Statement*: How do we structure an

agreement that the company can afford that brings sufficient value so the employees will ratify the agreement?

Morton Deutsch and Peter Coleman believe that reframing the conflict as a mutual problem to be cooperatively resolved by the parties is at the heart of a constructive dispute resolution process.[24] Also, they assert that conflict resolvers who are able to do this make it possible for the disputing parties to think creatively about the problem and the solution.[25] The following is an example of a reframed statement of the problems involved in a marital dispute.

> *Original Statement*: We have to decide who has custody, where the children will live, and how much time they will spend visiting the other parent.
>
> *Reframed Statement*: We have to decide how we are going to share our responsibilities as parents and what kind of time the children will spend in each of our homes.[26]

The potential use of reframing is unlimited. Here is an example of how it is done in a community dispute.

> *Original Statement*: Are we going to protect the unique quality of our community, or are we going to give in to the city's demands that we conform to the regulations that will in the end turn us into just one more yuppie neighborhood?
>
> *Reframed Statement*: How can we preserve the uniqueness of our community within the city's regulatory framework?[27]

Creating a New Relationship Paradigm

In Moore's view, when the focus of a dispute resolution process is on the "psychological condition of the disputants," the process becomes identified with the transformative style of mediation.[28] Robert Baruch Bush and Joseph Folger popularized transformative mediation in *The Promise of Mediation*.[29] The transformative approach to mediation had been employed in a wide variety of situations, ranging from community conflicts to international peacekeeping. Reframing plays an integral role in the transformative approach to mediation.

In the Baruch Bush-Folger model, the objectives of empowerment and "recognition" are not only central to the transformative approach, they are the most important objectives.[30] Recognition occurs when an individual acknowledges and empathizes with the situation and problems of others.[31]

Although Baruch Bush and Folger avoid using the term "reframing" and its variants, they invoke the concept when they say that mediators give a party to the mediation the opportunity to "translate" or "reinterpret" statements or behaviors of the other party into a more positive light. As they see it, acceptance of the reinterpreted statement or behavior is "one form of giving recognition."[32]

These authors also advert to reframing without saying so when they discuss the mediator's use of "nonimpositional summaries." These are summaries by the mediator of party statements that do not impose the mediator's own perspectives or views. In this transformative mediation model, the mediator looks for statements that can be used to evoke recognition in a party, for example, the attempted apology in the sexual harassment example mentioned above. Then the mediator "translates" the statement to the other party in terms that he or she can relate to, so that this person can respond and give recognition.

Jay Rothman, building on the works of others, developed the process of "reflexivity," which he uses to address conflicts rooted in the articulated frustrations and threats to the peoples' collective need for dignity, recognition, safety, control, purpose, and efficacy: what he calls "identity-based conflicts."[33] In Rothman's reflexive reframing, the parties articulate their needs and values that are being threatened or frustrated by the conflict and relations with adversaries. Through a dialogue with their adversary, they engage in open introspection and, it is hoped, thereby discover common ground—or what Rothman calls "resonance."[34]

Moving to a More Focused Phase

In both need-based and interest-based negotiations and conflict resolution, reframing assists in moving the parties away from positional bargaining to acknowledging their underlying needs and interests. Reframing is an integral part of a methodic approach to advancing the goals of the negotiation or dispute resolution process.

Mayer articulates a refinement on this theme which he calls "successive reframing." When complex conflicts or issues cannot easily be effectively reframed, the dispute resolution professional makes successive use of reframing. Successive reframing takes the parties through an "iterative refinement" of the statement of the problem, during which they gradually define and redefine the issues to be resolved until the statement resembles or suggests an acceptable solution.[35] The dispute resolution process begins with framing issues, then discovering interests, forming options, assessing options and then reframing the remaining options and issues. Although in this structure, the focus of the reframing effort is on the remaining options and issues, reframing may be used during any phase of the process.

Rothman approaches identity-based conflicts using his "ARIA" framework. This stands for "Antagonism through Adversarial Framing-Resonance through Reflexive Reframing-Invention through Inventing/Brainstorming Action through Agenda Setting." In the first stage, adversarial framing could be used to address what resources are at stake and what solutions are being sought. After this phase, reflexive reframing could focus on the "who" and "why" of the conflict—that is, the core identity needs of all parties. It is this step of the ARIA framework that can lead to Roth-

> *Various types of reframing have been developed to achieve different ends in the discussions with parties who are mediating.*

man's notion of resonance (or common ground), which in turn can foster harmony.[36] Rothman believes that in reflexive dialogue, the disputants reframe their perceptions and analyses of each other and their own identities. Thus, where blame is, mutual responsibility enters.[37] Reflexive reframing and its product, resonance, is the linchpin of Rothman's ARIA framework.

In these instances, dispute resolvers use reframing not only for a specific reason, such as to introduce a new perspective, clarify issues, or neutralize language; they use it as an integrated component and phase of the conflict resolution process itself.

Conclusion

At the heart of reframing verbal communication is the endeavor to get to the essential elements of the dispute and remove distractions, so that these elements may be understood and productively addressed. It is a versatile tool in negotiations and conflict resolution, with many applications. A greater appreciation of how the technique is used and its multiple applications could improve the results in the negotiation and conflict resolution setting. ∎

(Endnotes are on page 59)

tion. They are seriously concentrating on the dispute, often for the first time.

What is the result? With litigation, they are prepared, not always consciously, to resolve the dispute as much as they are prepared to try the case. In mediation, once the process begins, they want to participate to prove how good their case is, and they don't want to go through this a second time. They want to go back to business.

It is usually difficult to get everyone's attention on a dispute until some external event or authority requires it. The best argument for mediation, and for contractual provisions requiring mediation, is that it forces the event that will bring the parties and counsel together under circumstances favorable for settlement. A contractual requirement for mediation creates the necessary external pressure. It works.

Conclusion

Despite its successes, commercial mediation has not yet matured. Some corporate executives have taken the lead in encouraging the use of mediation, and trial lawyers in increasing numbers have often become experienced in mediation techniques. Governmental agencies have also signed on, since they have difficulty with binding ADR processes. This is forcing some members of the private sector to accept mediation in order to maintain their contracts with the government. Courts are also pushing almost all litigation into pre-trial mediation. All of this is energizing private mediation, which places control of the process, from start to finish, in the parties' hands.

Nothing is standing in the way of the movement toward mediation. But that movement will accelerate only when more transactional lawyers are educated about the process so that they can advise their clients at the inception of a transaction about ADR options, particularly mediation. This will lead to more mediation clauses and greater use of the mediation process. ∎

Reframing and Its Uses

ENDNOTES

[1] See Douglas H. Yarn, ed., *Dictionary of Conflict Resolution* (Jossey-Bass, 1999), p. 190. Yarn defines the term "reframing" in several ways. One of his definitions is "the technique of defining or redefining a particular concept or reality so as to give it a slightly different and more constructive perspective." He also defines reframing as a technique a mediator uses "to recast conflictual issues in neutral terms." Yarn points out that there are other terms for the technique, one of which is "framing." It is this term that Yarn defines broadly in terms of a frame of reference.

[2] William Ury, *Getting Past No* (Bantam Books, 1993), p. 78.

[3] *Id.* at 89-98.

[4] *Id.* at 92.

[5] Morton Deutsch & Peter T. Coleman, eds., *The Handbook of Conflict Resolution Theory and Practice* (Jossey-Bass, 2000), pp. 363-364.

[6] Bernard Mayer, *The Dynamics of Conflict Resolution, A Practitioner's Guide* (Jossey-Bass, 2000), p. 136.

[7] Ury, *supra* n. 2, at 80-89.

[8] *Id.* at 78.

[9] Christopher Moore, *The Mediation Process* (Jossey-Bass, 1996), p. 236.

[10] *Id.* at 217.

[11] Kimberlee K. Kovach, *Mediation Principles & Practice* (West Publishing Company, 1994), p. 108 .

[12] Dwight Golann, *Mediating Legal Disputes* (Little, Brown & Co., 1996), p. 71.

[13] Deutsch & Coleman, *supra* n. 5, at 506.

[14] Moore, *supra* n. 9, at 219.

[15] *Id.* at 215.

[16] *Id.* at 220.

[17] *Id.* at 222.

[18] Kovach, *supra* n. 11, at 108.

[19] Mayer, *supra* n. 6, at 134.

[20] *Id.*

[21] *Id.* at 135.

[22] *Id.*

[23] Moore, *supra* n. 9, at 242-243.

[24] Deutsch & Coleman, *supra* n. 5, at 31-32.

[25] *Id.* at 363.

[26] Mayer, *supra* n. 6, at 135.

[27] *Id.* at 136.

[28] Moore, *supra* n. 9, at 206.

[29] Robert A. Baruch Bush & Joseph P. Folger, *The Promise of Mediation: Responding to Conflict through Empowerment and Recognition* (Jossey-Bass, 1994).

[30] *Id.* at 84.

[31] *Id.* at 2.

[32] *Id.* at 125-126.

[33] Deutsch & Coleman, *supra* n. 5, at 445.

[34] Jay Rothman, *Resolving Identity-Based Conflict* (Jossey-Bass, 1997), p. 33.

[35] Mayer, *supra* n. 6, at 165-166.

[36] Rothman, *supra* n. 34, at 18-19.

[37] *Id.* at 49.

Frame Changes and the Management of Intractable Conflicts

DEBORAH SHMUELI

MICHAEL ELLIOTT

SANDA KAUFMAN

Frames play an important role in intractable conflict. As lenses through which disputants interpret conflicts, divergent frames limit the clarity of communication and the quality of information and encourage escalation. These frames, embedded in personal, social, and institutional roles, are often quite stable over time. Yet in some intractable conflicts, reframing interventions have contributed to tractability.

A riot is at bottom the language of the unheard.

> *Martin Luther King Jr.,*
> *American civil rights*
> *leader, 1967*

[Rioters] are lawbreakers, destroyers of constitutional rights and liberties and ultimately destroyers of a free America.

> *Lyndon Baines Johnson,*
> *American president, 1965*

Martin Luther King Jr. and Lyndon B. Johnson are describing the same event using different frames, or interpretive lenses, through which each individual views and makes sense of unfolding events. This article explores the nature of frames and the framing process and their consequences for intractable conflicts. We define the concept and review what is known about frames and their impact on conflict dynamics. We then discuss commonly recurring frames and offer some examples from practice. Finally, we examine the potential for, and implications of, reframing in conflict management.

The Nature of Frames

Disputants use frames both interpretively and strategically (Kaufman and Smith, 1999). As interpretive lenses, frames help us make sense of complex situations in ways internally consistent with our worldviews, giving meaning to events in the context of life experience and understandings. As strategic tools, frames help rationalize self-interest, persuade broader audiences, build coalitions, or promote preferred outcomes. Hence, Martin Luther King's and Lyndon Johnson's statements are two-edged: they result from both internal sense making and strategic intentions.

Frames as Sense Making

Most researchers cast frames as cognitive devices or shortcuts for making sense of complex information (Gray, 1997; Lewicki, Saunders, and Minton, 1999; Taylor, 2000; Goldratt, 1990; Kahneman and Tversky, 1979; Neale and Bazerman, 1985 and Vaughan and Seifert, 1992). We create frames to name situations in which we find ourselves, identify and interpret aspects that seem to us key in understanding the situation, and communicate that interpretation to others. These cognitive structures help reduce information overload and operate as models of reality that, by necessity, trade detail for clarity. Frames organize phenomena into coherent, understandable categories, giving meaning to some observed aspects while discounting others that appear irrelevant or counterintuitive. This selective simplification filters people's perceptions and defines their fields of vision. It can lead to sharply divergent interpretations of an event, as demonstrated by reactions to OJ Simpson's acquittal of murder charges, the riots that followed the beating of Rodney King, the reintroduction of wolves into Yellowstone National Park, the Israeli withdrawal from Gaza, or Elian Gonzalez's return to his father in Cuba.

As with models, what makes frames useful also makes them prone to error. Consider two environmental justice disputes. In Cleveland, neighborhood residents with a long experience of being misled by city and corporate officials frequently interpreted new officials' actions and communications as purposely misleading. In Chattanooga, residents opposed proposals to reduce hazardous air emissions because they viewed the proposed facilities as hazardous waste incinerators, with all the risks that label evokes. Residents and officials never met in dialogue because residents viewed officials as callous and manipulative, while officials considered

residents as having irrational perceptions and demands. These interpretive lenses remained remarkably stable for decades.

Frames as Communicative Strategy

From a communicative perspective, framing is strategic, aiming to persuade others to our point of view, gain advantage in negotiations, and rally like-minded people to our causes. Disputants usually formulate frames that are consistent with their interests. For example, when people tell each other what a conflict is about, they choose terms that advantage their side. Similarly, views about what is fair in a situation are often driven by parties' assessment of which fairness measures will benefit them most.

Identification with a group straddles the interpretive and communicative frame categories. We accept a group's framing of a situation either because we recognize it as our own or because we benefit from membership in that group. For example, those who see the environment as intrinsically valuable rather than utilitarian share a frame and join groups defending that point of view. Identifying with a group amounts to recognizing group members as similar to ourselves along important dimensions and ignoring differences on dimensions of lesser importance.

The Impact of Frames on the Intractability of Conflict

Frame divergence often contributes to the intractability of conflicts. Disputants differ not only in interests, beliefs, and values but also in how they perceive the situation at the conscious and preconscious levels. These differences engender divergent interpretations of events, paint parties into negative characters, yield mutually incompatible issues, and focus attention on specific outcomes that impede exploration of alternatives.

As conflicts become intractable, frame differences often exacerbate communication difficulties, polarize parties, and escalate strife. In turn, polarization is reflected in the parties' frames, feeding stakeholders' sense that they are in the right and should not compromise. Divergent frames are self-reinforcing because they filter parties' subsequent information intake and color interpretation and because disputants strategically communicate through these frames to strengthen their positions and persuade opponents.

Linked as they are to information processing, message patterns, linguistic cues, and socially constructed meanings, framing and reframing are vital to the communications underlying negotiations (Putnam and Holmer, 1992; Elliott, Gray, and Lewicki, 2003). While intractability is unlikely to hinge on frames alone—so reframing alone is unlikely to eliminate intractability—awareness of frame differences can help stakeholders and interveners in managing conflicts. Knowing the frames in use and how they were constructed helps us understand and influence conflict dynamics. With framing insight and through reframing, stakeholders may find new ways out of impasse. Interveners can use expressed frames to understand the situation and design interventions. At times, a stakeholder's awareness of others' frames can assist in mutual understanding and reframing of proposals in terms that might be more acceptable to the others.

Sources and Categories of Frames

What shapes our frames? Disputes are associated with complex and mutually reinforcing frames about oneself, the others, how decisions should be made, risks, and what information should apply to the situation. In what follows, we describe these frames, illustrating them with three cases of environmental conflict. We note that while framing is present in all conflicts, it is particularly prominent when conflicts are long lasting and alienation and escalation drive disputants apart. Therefore, frame insights may be especially useful for understanding, managing, and intervening in intractable conflicts. A brief overview of the three cases follows:

- Cleveland's air, while a far cry from steel mill days, remains in nonattainment for ozone and particulates. In 2000, the U.S. Environmental Protection Agency initiated and financed a pilot project to reduce air toxins through voluntary action in two Cleveland neighborhoods. The group convened to select toxin reduction projects was diverse in interests, knowledge, experience, and ability to implement changes.
- The Dudaim dispute revolved around the opposition of local and district authorities and nongovernmental organizations (NGOs) to the siting of a national landfill in Israel. The overriding concern was fear of environmental damage to future development and the quality of life in metropolitan Beer Sheva, located in the Negev, Israel's southern periphery.
- Since the 1970s, Chattanooga, Tennessee, has moved from a declining economy based in heavy industry to an urban renaissance rooted in

civic engagement and sustainable redevelopment. However, its remarkable success was not uniform. Old industry left a legacy of significant contamination in the Alton Park and Piney Woods neighborhoods. The mostly African American residents, who lived adjacent to environmentally degraded land and faced poverty and disinvestment, persistently conflicted with industries and public agencies over environmental cleanup.

Using these cases, we illustrate several types of frames, how they operate, and how they can contribute to intractability.

Identity Frames

Parties in conflict take on identities derived from the interplay between their self-conception and interests, and their group affiliations (Rothman, 1997). Challenges to one's sense of self trigger opposition and may even deflect attention from issues and toward protection of one's identity. Typical responses to identity challenges—ignoring information and perspectives that threaten it, reinforcing affiliations with like-minded groups, and negatively characterizing outsiders—impede subsequent agreements.

Opposing a national landfill that would primarily serve the Tel Aviv metropolis, Dudaim residents announced that they—the nation's southern periphery—were not the nation's garbage dump. History- and geography-driven identity frames set the country's periphery against its center. The Negev's hinterland status and its struggle against unfair central government decisions and for a more autonomous regional identity became central.

Identity frames are often salient and part of the polarized discourse in intractable conflicts. They rarely shift dramatically in the short range. Moreover, it seems that such a change, although helpful, is not necessary for managing conflicts or reaching agreements. Instead, interveners and stakeholders are better off focusing on reframing characterization frames.

Characterization Frames

Characterization frames are reductionist labels, associating positive or negative characteristics with individuals or groups. The strength of these frames lies in their being shared, so people can communicate them to others who understand them in the same way. In intractable conflicts, characterization frames may undermine opponents' legitimacy, cast doubt on their motivations, or exploit their sensitivity. Characterization and identity

frames are at times linked, strengthening one's own identity while justifying actions toward the other, as we frame opponents as our opposite.

In Cleveland, neighborhood residents initially held firm negative characterization frames: all businesses were polluters, and all city officials were unreliable and misleading. Although ostensibly engaged in a collaborative dialogue, residents took six months to reframe business representatives as project partners. In protracted conflicts, negative characterizations can linger for years, leading to few or no positive joint experiences, which reinforce the negativity.

Interveners can and should address and diffuse such mutual characterizations. As with identity frames, the more intractable the conflict is, the more salient mutual negative characterizations are. Unlike identity frames, characterization frames lend themselves to reframing with the assistance of a skillful intervener who creates opportunities for positive direct interactions.

Power, Social Control, and Conflict Management Frames

Intractable conflicts are often embedded in struggles to alter existing institutions or decision-making procedures. Disputants' conceptions of power (the basis on which social decisions are or should be made), social control (whether disputants feel they can influence their future as individuals or as part of groups), and conflict management (the legitimacy of particular approaches to resolving differences) are important in conflict dynamics. These frames shape disputants' assessment of which forms of power are legitimate and which are likely to advance their own position. The more intractable the conflict is, the more stakeholders are likely to interpret events as mutually exclusive power struggles, resulting in polarization. Traditional decision-making processes give way not to dialogic forms of dispute management (perceived as reinforcing existing power imbalances) but rather to adjudicatory, civil, or violent confrontations (legitimated by the perceived power imbalance).

Although altering such frames takes time, they are amenable to shifts as stakeholders experience the failure of unilateral, power-based approaches and the potential of collaborative ones. Constituting only 20 percent of the Chattanooga electorate, African Americans held little effective power. Alton Park residents in particular felt disempowered and alienated. Civil disobedience (for example, catching rats in public housing buildings and releasing them at city hall public hearings) constituted a relatively restrained resistance to the elite who "lived up on the mountain." This frame remained stable for

decades. Even as the city engaged in extensive, participatory visioning processes that initiated and legitimized civic dialogue, these residents remained on the outside. Only after a decade of witnessing the impact of dialogues on other neighborhoods did Alton Park residents claim a voice in the process. Passive resistance slowly gave way to active engagement.

Risk and Information Frames

Intractable disputes often revolve around events with risky or uncertain consequences. Disputants' risk frames may yield drastically diverging assessments of level and extent of risk. In addition, disputants develop frames regarding the reliability of information and its sources. Risk and information frames depend not only on the disputants' interests, but also on their expertise, level of exposure to, and familiarity with, the risk, the extent of dread evoked by the risk, its potential for catastrophic impacts, and whether the risk is a personal choice or imposed.

In Dudaim, national government agents and waste disposal operators were certain that the advanced technology planned for the landfill posed no environmental risk. Local residents, however, feared uncertain threats to their quality of life and environment, while accepting known risks from existing large petrochemical industries and an army camp. In Cleveland, a resident, while smoking, asked the transit authority to convert all its buses to low sulfur diesel because he worried about the health risks. He framed his exposure to bus exhaust while riding the bus to work (a social risk factor beyond his control) as riskier than his chain smoking (a personal risk factor under his control), although smoking is, by far, the higher health threat.

Loss and Gain Frames

People tend to perceive as more salient and work harder to prevent a loss than to capture a commensurate gain (Kahneman and Tversky, 1979). This perceptual asymmetry is manipulable, as the expected consequences of risky choices can often be framed as either a potential loss or a potential gain.

Alton Park residents strongly opposed efforts to clean up a heavily polluted neighborhood stream. For cleanup, tons of tar-based deposit removed from the stream had to be temporarily staged at the coking plant that had produced the waste. Residents feared the wastes might be left on-site indefinitely. This potential loss (though unlikely) seemed more salient to residents than the much larger potential health gains from cleanup. Residents blocked cleanup efforts for over a year and accepted the project

only after forming a community advisory panel to conduct detailed review of procedures.

Frame Analysis and Reframing as Conflict Management Tools

The frames described in the previous section are by no means immutable, though some are more amenable to change than others (Elliott, Gray, and Lewicki, 2003). Frame changes occur when new information or direct experience manages to overwhelm filters. For example, while characterization frames are often robust, positive interactions, particularly if they promote empathic consideration of the others' experiences, may lead to reframing.

How can a conflict practitioner analyze frames and promote reframing in an intractable conflict? We must note that while reframing is often used in dispute resolution practice, we know more about the nature of frames and their impact on conflicts than we do about practical strategies for managing frames, including their effectiveness and applicability to specific contexts. We do not know yet whether particular interventions yield specific changes in frames or how these changes, if they occur, alter the course of disputes. Yet research on frames has yielded insight into tentative strategies as well as underscored the importance of frames in conflict dynamics. With these caveats in mind, we explore the practical application of framing research to the management of conflict, particularly intractable conflict.

What Can the Understanding of Frames Contribute to Conflict Management?

Based on several observations, we argue that mediators and facilitators aware of disputants' frames are better able to intervene in intractable conflict. First, frames contribute to escalation and polarization, thereby impeding productive change. Second, when an intractable conflict changes course, it is usually accompanied by some changes in disputants' frames. Indeed, at times, parties have confronted their own frames and concluded these frames were limiting their ability to achieve their own goals. Third, in practice, the design of effective interventions requires understanding the prevailing frames held by disputants.

Within processes of reconciliation, negotiation, or joint problem solving, explicit management of frames and the framing process may lead to

shifts in the frames and their impact on conflict dynamics. Frame analysis and reframing—the purposive management of frames—aim to:

• Clarify or refresh perceptions of disputed issues, promote productive information exchange and listening, and expand the discussion framework. The Cleveland Air Toxics Pilot participants' view of themselves as passive recipients of the Environmental Protection Agency's enforcement services impeded progress until the parties saw themselves as actors with responsibilities and the means to contribute to change.
• Sharpen the parties' understanding of their own interests and how their frames and modes of action serve those interests. Alton Park residents, in focusing on potential losses and their sense of powerlessness to exercise positive control, worked against their own interests by opposing stream cleanup. Shifts in loss and gain, characterization, and power frames enabled residents to better understand their trade-offs and move from passive resistance to active problem solving.
• Identify issues that parties view differently and, based on this understanding of the divergent frames, identify opportunities for trade-offs. In the Dudaim dispute, the national authorities and waste disposal companies focused on the environmental and economic suitability of the chosen site, while local parties saw the struggle in terms of their peripheral status among national priorities. These frames are not necessarily incompatible but require creative, compensatory solutions to bridge differences effectively.
• Identify deep differences that cannot be bridged and design conflict reduction processes that do not violate these deep differences. In Chattanooga, after decades of litigation and open conflict over environmental impacts, Alton Park residents and a chemical manufacturing plant began a ten-year dialogue. During that time, stakeholders fundamentally differed in risk frames, yet they made considerable progress on increasing plant safety, improving its environmental integrity and aesthetics, and generating community benefits for the residents. Effective reduction in risk did not require a shared frame of what constituted risk.

How Are Frames Identified and Analyzed?

Stakeholders, interveners, and conveners can analyze frames to deepen their understanding of conflict dynamics. Frame analysis is conducted by holding in-depth conversations and interviews with stakeholders about their past experiences and current expectations from the conflict and each other. Such reflections offer lessons from the past (retrospective analysis)

and provide cues about difficulties to be expected in upcoming interactions (prospective analysis). By assessing the role of frames in the conflict, interveners can better design intervention strategies (Kaufman and Gray, 2003). Feeding back to parties information about their frames and exploring with them the meaning and impact of frames on their conflict dynamics may contribute to reframing and progress (Shmueli and Ben Gal, 2004).

In intractable conflicts, frames are often quite stable over time, even when specific individuals move in and out of the situation, because the frames are self-reinforcing, shared through community storytelling, and socially reinforced through community interaction. Yet in some intractable conflicts, frames change in time through interventions, and the frame shift helps reduce intractability (Elliott, Gray, and Lewicki, 2003). However, reframing is difficult: it requires parties to take on new perspectives and some risk. Therefore, reframing may have to be accompanied by changes in the dispute context creating incentives to consider new perspectives, or within careful, constructive dialogue focused on improving communication and building trust.

Among the strategies and techniques that use dialogue to foster reframing in intractable conflicts are these (Elliott, Gray, and Lewicki, 2003):

• Deescalatory processes, including listening projects, study circles, and some forms of mediation, seek to reduce tension through forums promoting effective communication around a set of limited objectives. Rather than seeking consensus, these forums focus on reducing escalatory cycles by empowering disputants to communicate directly without needing to defend current positions. Study circles structure small-group interactions using materials the disputants themselves design. Listening projects send disputants to interview other stakeholders about their perspectives and concerns. Such processes build on the premise that active listening around specific issues helps disputing parties move more deeply into the causes of their conflicts and builds understanding, trust, and mutual respect.

• Perspective-taking processes, including imaging of identities and characterizations, narrative forums, and story projects, help disputants understand the conflict and its dynamics from other disputants' perspectives. These approaches are particularly geared toward understanding identity and characterization frames. Through explicit identification of critical identities, role playing, the creation and listening to others' stories, and acknowledgment of past injustices, these approaches enable disputants to see the validity and credibility of other perspectives and examine the interplay between their own and other disputants' frames.

• Identification of commonalities, including search for common ground and visioning and search processes, enable reframing around a smaller number of issues than those contributing to intractability. Common ground processes are used in highly divisive issues (such as abortion rights and ethnic disputes) and explore areas of agreement and possible joint action between parties who normally focus on their differences. Search processes seek to identify desired futures to shift the focus from a short- to a long-term perspective. By shifting attention away from areas of hardened differences and onto areas of potential overlap, these processes seek to break down rigid characterization frames and promote actions that open communication and build relationships across old divides.

Other intervention approaches enhance the desirability of alternative options to parties with divergent frames. For disputants to examine options from the perspective of other parties, they must understand the others' frames and view options from other perspectives. Seeking to reframe perceptions of losses as gains can enhance the openness and creativity of disputants.

Conclusion

Frames play an important role in intractable conflict. As lenses through which disputants interpret conflicts, divergent frames limit the clarity of communication and the quality of information, and they encourage escalation. These frames, embedded in personal, social, and institutional roles, are often stable over time, even across the ebb and flow of many dispute episodes. Yet in some intractable conflicts, frame-based interventions have led to reframing that has contributed to tractability. In addition, frame analysis can inform disputant strategies and intervention design. Research is needed to explore which energizing events move disputants to reexamine the basis of their dispute. Simpler ways must be devised for eliciting framing information during interventions to make this tool more accessible to interveners.

References

Elliott, M., Gray, B., and Lewicki, R. "Lessons Learned About the Framing of Intractable Environmental Conflicts." In R. Lewicki, B. Gray, and M. Elliott (eds.), *Making Sense of Intractable Environmental Conflicts: Concepts and Cases.* Washington, D.C.: Island Press, 2003.

Gray, B. "Framing and Reframing of Intractable Environmental Disputes." In R. Lewicki, R. Bies, and B. Sheppard (eds.), *Research on Negotiation in Organizations, 6.* Greenwich, Conn.: JAI Press, 1997.

Kahneman, D., and Tversky, A. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica,* 1979, *47,* 263–289.

Kaufman, S., and Gray, B. "Using Retrospective and Prospective Frame Elicitation to Evaluate Environmental Disputes." In L. Bingham and R. O'Leary (eds.), *Evaluating Environmental Dispute Resolution.* Washington, D.C.: Resources for the Future, 2003.

Kaufman, S., and Smith, J. "Framing and Reframing in Land Use Change Conflicts." *Journal of Architectural and Planning Research,* 1999, *16* (2), 164–180.

Lewicki, R., Saunders, D., and Minton, J. *Negotiation.* New York: McGraw-Hill, 1999.

Neale, M. A., and Bazerman, M. H. "The Effects of Framing and Negotiator Overconfidence on Bargaining Behaviors and Outcomes." *Academy of Management Journal,* 1985, *28,* 34–49.

Putnam, L., and Holmer, M. "Framing, Reframing, and Issue Development." In L. Putnam and M. E. Roloff (eds.), *Communication and Negotiation.* Thousand Oaks, Calif.: Sage, 1992.

Rothman, J. *Resolving Identity-Based Conflict in Nations, Organizations, and Communities.* San Francisco: Jossey-Bass, 1997.

Shmueli, D., and Ben Gal, M. "The Potential of Framing in Managing and Resolving Environmental Conflict." In E. Feitelson, G. de Roo, and D. Miller (eds.), *Advancing Sustainability at the Sub-National Level.* Aldershot, England: Ashgate Press, 2004.

Taylor, D. E. "The Rise of the Environmental Justice Paradigm: Injustice Framing and the Social Construction of Environmental Discourses." *American Behavioral Scientist,* 2000, *43* (4), 508–580.

Vaughan, E., and Seifert, M. "Variability in the Framing of Risk Issues." *Journal of Social Issues,* 1992, *48* (4), 119–135.

Deborah Shmueli is senior lecturer in the Department of Geography and Environmental Studies at the University of Haifa.

Michael Elliott is associate professor of planning and public policy at the Georgia Institute of Technology's City and Regional Planning Program, associate with the Consortium on Negotiation and Conflict Resolution, and principal with the Southeast Negotiation Network.

Sanda Kaufman is professor of planning and public administration at Cleveland State University's Levin College of Urban Affairs.

Morally Reframed Arguments Can Affect Support for Political Candidates

Jan G. Voelkel[1] and Matthew Feinberg[2]

Abstract

Moral reframing involves crafting persuasive arguments that appeal to the targets' moral values but argue in favor of something they would typically oppose. Applying this technique to one of the most politically polarizing events—political campaigns—we hypothesized that messages criticizing one's preferred political candidate that also appeal to that person's moral values can decrease support for the candidate. We tested this claim in the context of the 2016 American presidential election. In Study 1, conservatives reading a message opposing Donald Trump grounded in a more conservative value (loyalty) supported him less than conservatives reading a message grounded in more liberal concerns (fairness). In Study 2, liberals reading a message opposing Hillary Clinton appealing to fairness values were less supportive of Clinton than liberals in a loyalty-argument condition. These results highlight how moral reframing can be used to overcome the rigid stances partisans often hold and help develop political acceptance.

Keywords

moral psychology, political psychology, influence, attitudes

Political elections provide the general populace with a choice between two (or more) candidates who have contrasting viewpoints on how best to serve the people and their needs. These viewpoints, and the candidates who endorse them, often fall into opposing sides of the political spectrum, with one candidate representing a more liberal perspective and the other representing a more conservative perspective. Generally, people's support for one candidate or the other reflects whether they identify with the liberal or conservative perspective that each candidate embodies (e.g., gallup.com, n.d.).

Despite these ideological allegiances, candidates, their campaigns, and everyday supporters invest substantial time and resources in hopes of persuading those who endorse the other candidate that he or she is the wrong person for the job. Even though this process is commonplace and exorbitant amounts of money are devoted to it (Cummings, 2008), it is largely unknown whether such attempts at persuasion are ever effective and, if they are, what types of arguments have the greatest impact.

In the present research, we explore the possibility that certain types of moral arguments can be an effective strategy for persuading liberals and conservatives to be less attached to the candidate who represents their party and perspective. Specifically, we examine the effectiveness of a technique called "moral reframing" (Feinberg & Willer, 2013, 2015) in the context of the United States 2016 presidential election.

Morality and Political Attitudes

Morality matters for political attitudes (Janoff-Bulman, Sheikh, & Baldacci, 2008; Morgan, Skitka, & Wisneski, 2010).

Liberals and conservatives possess different moral worldviews, and such differences help explain many of the contrasting stances the two sides take (Caprara, Schwartz, Capanna, Vecchione, & Barbaranelli, 2006; Graham, Haidt, & Nosek, 2009; Thorisdottir, Jost, Liviatan, & Shrout, 2007). Recently, researchers mapped the moral domain and found evidence for five moral foundations that form the basis of moral beliefs and judgments (Graham et al., 2011; Haidt & Josephs, 2004). The harm/care foundation is concerned with other's suffering and the need to prevent and alleviate such suffering. The fairness/cheating foundation relates to justice, equality, and discrimination. The loyalty/betrayal foundation emphasizes the importance of one's in-group and prioritizing that in-group. The authority/subversion foundation deals with respect for higher ranked individuals as well as adherence to tradition. Finally, the sanctity/degradation foundation is concerned with sacredness and purity and avoiding disgust-evoking behaviors (Haidt, 2007, 2012; Haidt & Joseph, 2004). Research has, in turn, found that compared to conservatives, liberals more strongly endorse the harm/care and the fairness/cheating foundations,

---

[1] Department of Social Psychology, Tilburg University, Tilburg, the Netherlands
[2] Rotman School of Management, University of Toronto, Toronto, Ontario, Canada

Corresponding Author:
Jan G. Voelkel, Tilburg University, Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, the Netherlands.
Emails: j.g.volkel@tilburguniversity.edu; jangvoelkel.research@gmail.com

while conservatives more strongly endorse the loyalty/betrayal, authority/subversion, and sanctity/degradation foundations (Graham et al., 2009; Haidt & Graham, 2007).

Moral Reframing and Candidate Arguments

Building on this understanding of the moral divide between liberals and conservatives, recent research has shown that it is possible to capitalize on these distinctions for purposes of political persuasion and coalition formation by using "moral reframing" (Day, Fiske, Downing, & Trail, 2014; Feinberg & Willer, 2013, 2015; Kidwell, Farmer, & Hardesty, 2013; Wolsko, Ariceaga, & Seiden, 2016). Moral reframing involves framing arguments that favor one's own political stance but grounding these arguments in moral terms that appeal to the moral values of those on the other side of the political spectrum. In this research, while liberals were unmoved by arguments in favor of conservative policies grounded in the more conservative moral foundations, their support for the conservative positions increased after reading messages grounded in the more liberal foundations, and this research also demonstrated the reverse when it comes to liberals persuading conservative targets (Feinberg & Willer, 2013, 2015; Kidwell et al., 2013; Wolsko et al., 2016; cf. Day et al., 2014).

Although this past research has shown that moral reframing can be an effective strategy for persuading those on the other side of the political spectrum to be more supportive of policies they would typically oppose, no research has explored the effectiveness of moral reframing in one of the most contentious, but fundamental, political domains—political campaigns. Might moral reframing be an effective means for affecting support for political candidates? We expected that it would, because moral evaluations are particularly relevant for person perception and impression formation overall (Goodwin, Piazza, & Rozin, 2014), and are especially relevant when making judgments about powerful figures and political candidates (Chen, Jing, & Lee, 2012; Skitka & Bauman, 2008; Trevino, Hartman, & Brown, 2000).

Additionally, moral reframing research has primarily focused on the effectiveness of morally reframed messages in support of a stance and has largely not explored whether this technique would work when arguments are made in opposition to a stance. Even so, understanding moral reframing's effectiveness in decreasing a target's support is particularly important, considering how much political rhetoric aims to decrease support for a policy or a political candidate. We predicted that the same underlying processes will apply regardless of whether a morally reframed message is in favor or in opposition to a stance; as long as, the argument itself is framed in a manner that appeals directly to the moral values of the targets, then those targets should be responsive to it because it fits with their morality.

The Present Research

We tested our predictions by examining the effectiveness of morally reframed messages in the context of the U.S.

presidential election campaign of 2016, presenting participants with short campaign messages in opposition to either Donald Trump (Study 1) or Hillary Clinton (Study 2). In each study, these messages were framed in terms of either a moral value endorsed at higher levels by conservatives (i.e., loyalty) or a moral value endorsed at higher levels by liberals (i.e., fairness). We expected that conservatives would become less supportive of Donald Trump after reading an oppositional message grounded in loyalty values than after reading a message grounded in fairness values. On the other hand, we expected liberals would become less supportive of Hillary Clinton after reading an oppositional message grounded in fairness values than after reading a message grounded in loyalty values.

We did not make any specific predictions regarding how liberals would respond to the anti-Trump messages and how conservatives would respond to the different messages in opposition to Clinton. Although the anti-Trump messages framed in more liberal moral terms might resonate with liberals and the anti-Clinton messages framed in more conservative moral terms might resonate with conservatives, these arguments may still be ineffective because they were aiming to persuade targets to take on a position that, likely, they already held (cf. Day et al., 2014).

Study 1

In the first study, we presented participants with arguments opposing Donald Trump that were framed in terms of either fairness or loyalty moral concerns. We hypothesized that conservatives in the loyalty argument condition would support Trump less than conservatives in the fairness argument condition, but the moderate and liberal participants would likely be unaffected by our manipulation. We measured support for Donald Trump, our dependent variable, with both attitudes (warmth and acceptance as president) and behavioral intentions (likelihood to vote for Trump) and tested whether the effect of experimental condition on the likelihood to vote for him might be mediated by the attitudes measures.

Method

Participants

Based on the past research on moral reframing (e.g., Feinberg & Willer, 2013, 2015), we expected a small effect size (specifically, a $R^2\Delta$ of approximately .02 to .03 as a result of including the interaction of political ideology and experimental condition into the regression equation). In order to have enough statistical power, therefore, we estimated a sample size of around 400 participants in each study would be required. In Study 1, 404 participants recruited from the Amazon Mechanical Turk website completed the study. Participants were excluded if they had missing values ($n = 3$) or if they failed an attention check ($n = 4$). Thus, the final sample size consisted of 397 participants (189 male, 207 female, 1 other; $M_{age} = 37.33$, $SD = 12.94$). Participants took part in this study on August 28, 2016, 72 days

prior to the 2016 presidential election and were given a small payment for their participation.

Procedure

Participants learned they would be presented with some information about a candidate for the 2016 presidential election and be asked questions afterward. Participants were then randomly assigned to one of two conditions: the loyalty or fairness argument condition. Both conditions involved presenting participants with a short message arguing against Donald Trump, modeled after actual campaign advertisements. The loyalty message was written so that it would appeal to the loyalty/betrayal moral foundation, incorporating words and phrases representative of that foundation (cf. Graham et al., 2009). For instance, the loyalty message argued that Trump "has repeatedly behaved disloyally towards our country to serve his own interests" and that "during the Vietnam War, he dodged the draft to follow his father into the development business" (for full text, see Supplemental Material). The fairness argument, in contrast, appealed to the fairness/cheating moral foundation and used words and phrases representative of that foundation. For instance, it argued that Trump "openly discriminates against Muslims threatening their rights to be treated with fairness and equality" and that "his unfair statements are a breeding ground for prejudice" (for full text, see Supplemental Material). Each message was accompanied by a picture of Donald Trump further highlighting the corresponding moral value, showing him either next to American soldiers in action (loyalty argument condition) or next to Muslims demonstrating against terrorism (fairness argument condition).

Following the campaign message, participants were asked to summarize the message they just read, which served as an attention check. Two raters coded whether participants' answers to the attention check indicated that the participants actually read the arguments. The interrater reliability was high ($\varphi = .70$). We excluded only those participants for which both coders rated the summary as inadequate. Afterward, participants completed 3 measures relating to Donald Trump. *Warmth* was measured with the item: "How warm or cold do you feel toward Donald Trump?" answered on a scale from 0 (*very cold*) to 100 (*very warm*). *Acceptance as President* was measured with the item: "How easy or hard would it be for you to accept Donald Trump as the President of the United States?", answered on a scale from 0 (*very easy*) to 100 (*very hard*). Finally, *Likelihood to Vote* was measured with the item: "In the upcoming 2016 presidential election, how likely are you to vote for Donald Trump for president?," answered on a scale from 0 (*very unlikely*) to 100 (*very likely*). The initial position of the slider for all 3 items was at the midpoint of the scales. Finally, participants completed a demographic questionnaire which included a measure of political ideology ("Generally speaking, do you usually think of yourself as conservative, moderate, or liberal?") with three response categories (conservative, moderate, and liberal).

Analysis Strategy

We conducted separate multiple regression analyses for the three dependent variables. A dummy variable for moral argument condition (fairness argument as reference group), two dummy variables for political ideology (conservatives as reference group), and the interaction terms of condition and ideology were included as independent variables.

Although we expected different effects of the moral argument condition for the different ideology groups (implying an interaction effect), our main focus was a priori on the simple slopes analyses. To ensure the robustness of our results, we included several robustness checks that consistently supported our results (for details, see Supplemental Material). In addition, we conducted a moderated mediation analysis using Model 8 of Hayes' Process macro (Hayes, 2013). We included experimental condition as the independent variable, ideology as the moderator, *warmth* and *acceptance as president* as mediators, and *likelihood to vote for Trump* as the dependent variable. A bias-corrected bootstrap estimation approach with 5,000 samples was used to estimate the indirect effects.

Results

Means and standard deviations of the dependent variables for each condition by ideology group are presented in Table 1.

Warmth

The regression analysis showed a significant interaction effect, $\Delta R^2 = 0.01$, $F(2, 391) = 3.14$, $p = .044$. Simple-slopes analyses indicated that, as expected, conservative participants perceived Trump as less warm in the loyalty argument condition than in the fairness argument condition, $b = -13.82$, $t(391) = -2.53$, $p = .012$, 95% confidence interval (CI) = $[-24.58, -3.06]$. There was no significant effect of the moral argument condition for either moderates, $b = -1.43$, $t(391) = -0.34$, $p = .736$, 95% CI $[-9.75, 6.90]$, or liberals, $b = 2.90$, $t(391) = 0.75$, $p = .453$, 95% CI $[-4.69, 10.50]$.

Acceptance as President

The regression analyses showed a significant interaction effect, $\Delta R^2 = .01$, $F(2, 391) = 3.48$, $p = .032$. Simple-slopes analyses indicated that, as expected, conservative participants accepted Trump less as president in the loyalty argument condition than in the fairness argument condition, $b = -15.39$, $t(391) = -2.20$, $p = .028$, 95% CI $[-29.14, -1.65]$. There was no significant effect of the moral argument condition for either moderates, $b = 1.13$, $t(391) = 0.21$, $p = .835$, 95% CI $[-9.51, 11.76]$, or liberals, $b = 7.09$, $t(391) = 1.44$, $p = .152$, 95% CI $[-2.62, 16.80]$.

Likelihood to Vote for Trump

The interaction effect was significant, $\Delta R^2 = .02$, $F(2, 391) = 4.84$, $p = .008$. Simple-slopes analyses indicated that, as

Table 1. Results of Study 1: Means (*SD*s, *n*) for Argument Condition × Participants' Ideology.

	Ideology		
Condition	Conservative	Moderate	Liberal
(a) Warmth			
Fairness argument	61.04 (31.56, 45)	31.87 (30.50, 67)	5.10 (12.63, 83)
Loyalty argument	47.23 (32.65, 40)	30.44 (30.54, 75)	8.00 (14.86, 87)
(b) Acceptance as president			
Fairness argument	65.84 (32.94, 45)	34.55 (34.00, 67)	11.51 (25.53, 83)
Loyalty argument	50.45 (36.74, 40)	35.68 (34.77, 75)	18.60 (31.49, 87)
(c) Likelihood to vote			
Fairness argument	74.62 (31.72, 45)	31.58 (37.56, 67)	2.06 (8.88, 83)
Loyalty argument	55.75 (39.82, 40)	31.13 (38.31, 75)	7.71 (19.64, 87)

Note. The acceptance as president measure was recoded so that higher values indicate that participants were more willing to accept Trump as president.

expected, conservative participants were less likely to vote for Trump in the loyalty argument condition than in the fairness argument condition, $b = -18.87$, $t(391) = -2.91$, $p = .004$, 95% CI $[-31.61, -6.14]$. There was no significant effect of the moral argument condition for either moderates, $b = -0.45$, $t(391) = -0.09$, $p = .929$, 95% CI $[-10.30, 9.40]$, or liberals, $b = 5.65$, $t(391) = 1.24$, $p = .217$, 95% CI $[-3.34, 14.65]$. These findings are illustrated in Figure 1.

Moderated Mediation Analyses

The results of the moderated mediation analysis were consistent with our hypotheses. For conservatives, the effect of experimental condition on the likelihood to vote for Trump was mediated by warmth, $b = -12.15$, $SE = 6.26$, 95% CI $[-25.07, -0.56]$, and by acceptance as president, $b = -1.73$, $SE = 1.07$, 95% CI $[-4.69, -0.22]$, and the direct effect of experimental condition on likelihood to vote for Trump was not significant for conservatives, $b = -4.99$, $t(389) = -1.36$, $p = .173$, 95% CI $[-12.19, 2.21]$. For moderates, there was no indirect effect of experimental condition on the likelihood to vote for Trump, for warmth: $b = -1.25$, $SE = 4.54$, 95% CI $[-10.19, 7.75]$, or for acceptance as president: $b = 0.13$, $SE = 0.70$, 95% CI $[-1.27, 1.61]$. The direct effect of experimental condition on likelihood to vote for Trump was also not significant for moderates, $b = 0.68$, $t(389) = 0.24$, $p = .810$, 95% CI $[-4.85, 6.20]$. Likewise, for liberals, there was no indirect effect of experimental condition on the likelihood to vote for Trump, for warmth: $b = 2.55$, $SE = 1.88$, 95% CI $[-1.15, 6.32]$, or for acceptance as president: $b = 0.80$, $SE = 0.56$, 95% CI $[-0.04, 2.27]$, and the direct effect of experimental condition on likelihood to vote for Trump was not significant for liberals, $b = 2.30$, $t(389) = 0.90$, $p = .371$, 95% CI $[-2.75, 7.36]$.

Discussion

We found causal evidence that, compared to arguments in opposition to Donald Trump grounded in fairness concerns, arguments opposing Trump that appealed to the more conservative value of loyalty were more effective in causing
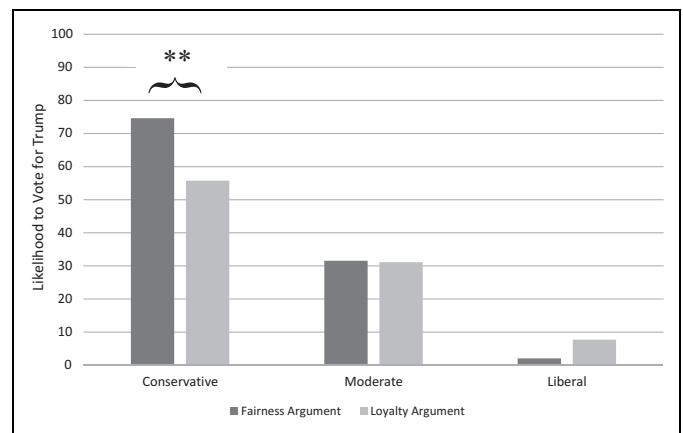

Figure 1. Likelihood to vote for Trump depending on argument condition and participants' ideology. $*p < .05$, $**p < .01$.

conservative participants to feel colder toward Trump, to accept him less as president, and, most importantly, to be less likely to vote for him. Further, the results suggest that the effect of moral argument condition on the likelihood to vote for Trump was mediated by perceived warmth and acceptance as president for conservatives. We did not find convincing evidence that the moral argument condition affected the support of moderates or liberals for Donald Trump.

Study 2

In Study 2, we aimed to conceptually replicate Study 1 with Hillary Clinton as the target instead of Donald Trump. That is, we presented participants with arguments opposing Hillary Clinton's candidacy that were framed in terms of either fairness or loyalty moral concerns. We hypothesized that liberals in the fairness argument condition would support Clinton less than liberals in the loyalty argument condition, while the manipulation would not affect the moderates or conservatives. We measured support for Hillary Clinton with the same measures as in Study 1 and tested whether the effect of experimental condition on the likelihood to vote for her would be mediated by the attitudes measures.

Table 2. Results of Study 2: Means (SDs, n) for Argument Condition × Participants' Ideology.

Condition	Ideology		
	Conservative	Moderate	Liberal
(a) Warmth			
Fairness argument	10.59 (23.21, 37)	27.45 (28.10, 74)	42.04 (27.17, 84)
Loyalty argument	9.08 (20.61, 39)	29.45 (29.03, 71)	54.59 (27.24, 87)
(b) Acceptance as president			
Fairness argument	25.73 (37.81, 37)	34.92 (32.26, 74)	60.95 (30.80, 84)
Loyalty argument	10.31 (19.42, 39)	35.25 (33.16, 71)	61.07 (33.77, 87)
(c) Likelihood to vote			
Fairness Argument	10.70 (26.50, 37)	33.88 (38.81, 74)	63.26 (38.32, 84)
Loyalty argument	10.41 (24.87, 39)	36.20 (40.71, 71)	75.98 (31.11, 87)

Note. The acceptance as president measure was recoded so that higher values indicate that participants were more willing to accept Clinton as president.

Method

Participants

Four hundred and eight participants recruited from the Amazon Mechanical Turk website completed the study. Participants were excluded if they had missing values ($n = 3$) or if they failed the attention check ($n = 13$). Thus, the final sample size consisted of 392 participants (172 males, 218 females, 1 agender, 1 genderqueer; $M_{age} = 36.86$, $SD = 12.24$). Participants took part in this study on September 2, 2016, 67 days prior to the 2016 presidential election and were given a small payment for their participation.

Procedure

The procedure paralleled that of Study 1, except the target of the message this time was Hillary Clinton instead of Donald Trump. Accordingly, we formulated messages in opposition to Clinton grounded in either loyalty or fairness values. For instance, the loyalty message argued that Clinton "is willing to risk the standing of our nation to achieve her own goals" and that "she failed our ambassador and soldiers in Benghazi" (for full text, see Supplemental Material). The fairness argument, in contrast, argued that "while so many Americans have suffered during the recent recession that the Wall Street Banks helped cause, Clinton has accepted millions of dollars from them in exchange for giving a few speeches" and that "Clinton is willing to sacrifice fairness and equality to achieve her own goals" (for full text, see Supplemental Material). The loyalty argument was accompanied by a picture showing Hillary Clinton next to an open envelope with an email symbol inside. The fairness argument was accompanied by a picture showing Hillary Clinton next to a Wall Street sign.

Following the campaign message, participants were asked to summarize the message they just read. As in Study 1, two raters coded whether participants' answers to the attention check indicated that the participants actually read the arguments. The interrater reliability was high ($\varphi = .89$). We excluded only those participants for which both coders rated the summary as inadequate. Afterward, they completed the same three

measures that were used in Study 1 regarding Hillary Clinton (warmth, acceptance as president, likelihood to vote). At the end of the study, participants completed a demographic questionnaire which included the same measure of political ideology as used in Study 1.

Analysis Strategy

We used the same analysis strategy as in Study 1 except that this time the loyalty condition was used as reference category for the moral argument manipulation. Again, we used several robustness checks that consistently supported our results (for details, see Supplemental Material).

Results

Means and SDs of the dependent variables for each condition by ideology group are presented in Table 2.

Warmth

The regression analysis showed a marginally significant interaction effect, $\Delta R^2 = .01$, $F(2, 386) = 2.43$, $p = .090$. Simple-slopes analyses indicated that, as expected, liberal participants perceived Clinton as less warm in the fairness argument condition than in the loyalty argument condition, $b = -12.55$, $t(386) = -3.06$, $p = .002$, 95% CI [-20.61, -4.49]. There was no significant effect of the moral argument condition for either moderates, $b = -2.00$, $t(386) = -0.45$, $p = .653$, 95% CI [-10.76, 6.75], or conservatives, $b = 1.52$, $t(386) = 0.25$, $p = .805$, 95% CI [-10.57, 13.61].

Acceptance as President

The interaction effect was not significant, $\Delta R^2 = 0.01$, $F(2, 386) = 1.83$, $p = .162$. In addition, simple effects analysis did not provide support for our hypothesis: Liberals in the fairness argument condition did not accept Clinton significantly less as president than liberals in the loyalty argument condition, $b = -0.12$, $t(386) = -0.02$, $p = .981$, 95% CI [-9.74, 9.51]. Additionally, there was no significant effect of the moral argument
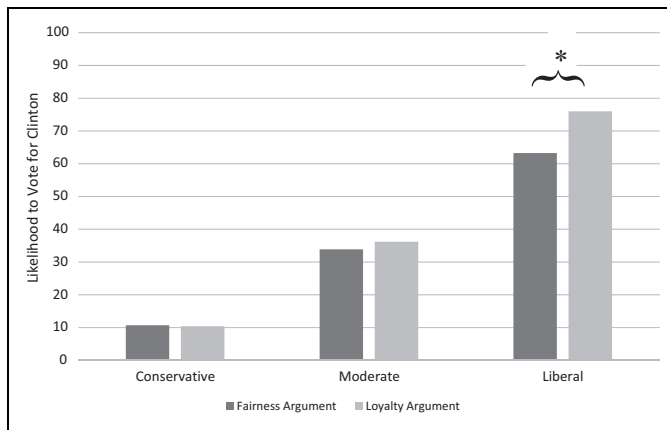
Figure 2. Likelihood to vote for Clinton depending on argument condition and participants' ideology. *p < .05, **p < .01.

condition for moderates, $b = -0.33$, $t(386) = -0.06$, $p = .950$, 95% CI [-10.79, 10.12], but there was some evidence that conservatives in the fairness argument condition accepted Clinton more as president than conservatives in the loyalty argument condition, $b = 15.42$, $t(386) = 2.10$, $p = .036$, 95% CI [0.98, 29.87].

Likelihood to Vote for Clinton

The interaction effect was not significant, $\Delta R^2 = 0.00$, $F(2, 386) = 1.27$, $p = .282$. However, simple-slopes analyses indicated that, as expected, liberal participants were less likely to vote for Clinton in the fairness argument condition than in the loyalty argument condition, $b = -12.72$, $t(386) = -2.36$, $p = .019$, 95% CI [-23.32, -2.11]. There was no significant effect of condition for either moderates, $b = -2.32$, $t(386) = -0.40$, $p = .692$, 95% CI [-13.83, 9.20], or conservatives, $b = 0.29$, $t(386) = 0.04$, $p = .971$, 95% CI [-15.62, 16.20]. These findings are illustrated in Figure 2.

Moderated Mediation Analyses

Although the interaction effects above were either marginal or not significant, we still chose to conduct moderated mediation analyses in line with Rucker, Preacher, Tormala, and Petty's (2011) argument that a significant total effect is not a requirement for a significant indirect effect to occur. However, it should be emphasized that such mediation analyses of nonsignificant effects should be interpreted cautiously. For liberals, the effect of experimental condition on the likelihood to vote for Clinton was mediated by warmth, $b = -10.80$, standard error (SE) = 3.76, 95% CI [-18.36, -3.84], but not by acceptance as president, $b = -0.02$, $SE = 1.00$, 95% CI [-2.18, 1.84], and the direct effect of experimental condition on likelihood to vote for Clinton was not significant for liberals, $b = -1.89$, $t(384) = -0.53$, $p = .596$, 95% CI [-8.90, 5.12]. For moderates, there was no indirect effect of experimental condition on the likelihood to vote for Clinton, for warmth: $b = -1.73$, $SE = 4.03$, 95% CI [-9.74, 6.20], or for

acceptance as president: $b = -0.06$, $SE = 1.09$, 95% CI [-2.14, 2.31], and the direct effect of experimental condition on likelihood to vote for Clinton was not significant for moderates, $b = -0.53$, $t(384) = -0.14$, $p = .890$, 95% CI [-8.02, 6.96]. For conservatives, there was an unexpected indirect effect of experimental condition on the likelihood to vote for Clinton via acceptance as president, $b = 2.99$, $SE = 1.51$, 95% CI [0.72, 6.85], but not for warmth: $b = 1.31$, $SE = 4.37$, 95% CI [-7.21, 10.05]. The direct effect of experimental condition on likelihood to vote for Clinton was not significant for conservatives, $b = -4.01$, $t(384) = -0.76$, $p = .450$, 95% CI [-14.43, 6.41].

Discussion

The results of Study 2, though in line with our hypotheses, were more complex than the results of Study 1. Although the results of the simple effects analyses were generally consistent with the predictions of a moral reframing account, the predicted interaction effects were nonsignificant for two of our dependent variables and marginally significant for the third dependent variable. Interestingly, we found some evidence for a moral reframing effect for conservatives, which, though not directly hypothesized, is in the direction predicted by a moral reframing account such that conservatives were more persuaded by appeals grounded in the more conservative moral foundation of loyalty.

General Discussion

Across two studies using the two major candidates from the 2016 U.S. presidential election as targets, we found evidence that moral reframing can be an effective strategy for persuading the electorate about political candidates (for a discussion of the differences between Study 1 and Study 2 see below). As such, the present inquiry extends past research on moral reframing in important ways. Although it has been shown that moral reframing can increase the support of liberals and conservatives for policies that they would usually oppose (e.g., Feinberg & Willer, 2013, 2015), the present research provides the first evidence that moral reframing is also an effective strategy to decrease the attachment of liberals and conservatives to the political candidate of the party they typically support.

Furthermore, the current findings illustrate that despite the fundamental moral differences separating liberals and conservatives (e.g., Graham et al., 2009; Jost, Glaser, Kruglanski, & Sulloway, 2003; Lakoff, 2002), there are ways that people across the ideological spectrum can make their stance understandable to a person from the other side. Although much research has outlined the enormous difficulties involved in fostering productive conversations and collaborations between liberals and conservatives (e.g., Brandt, Reyna, Chamber, Crawford, & Wetherell, 2014; Toner, Leary, Asher, & Jongman-Sereno, 2013), the current research highlights a technique where supporters of political candidates are responsive to

criticism about their favored candidate, and as a result, decreases the distance between liberals and conservatives.

The effectiveness of moral reframing raises the question of whether campaigns, pundits, and everyday people actually employ this technique to affect people's opinion about political candidates. In a first attempt to investigate this question, we asked liberal Clinton supporters and conservative Trump supporters to write arguments aimed at convincing those who endorse the other candidate as to why they should instead oppose him or her. In addition, we investigated the content of YouTube videos opposing Hillary Clinton or Donald Trump. Our results suggested that conservatives used morally reframed arguments more than liberals (cf. Haidt, 2012). However, this evidence should be viewed as only preliminary and fodder for future research (for more information, see Supplemental Material).

Overall, the present research had several important limitations. First, the support for a moral reframing effect was generally stronger in Study 1 than in Study 2. This difference could be driven by a number of factors. For instance, the content of the fairness and loyalty messages in the two studies was different, opening up the possibility that the quality or intensity of the arguments may have differed across studies. Another possibility could be that Trump's candidacy might be more strongly associated with issues of morality than Clinton's candidacy. Therefore, it might be easier to stimulate people's moral intuitions relating to Trump. Future research could potentially address this question by examining archival data (e.g., open-ended responses collected in polls), testing the extent to which beliefs about Trump and Clinton reflect individuals' core moral values and convictions (cf. Skitka & Bauman, 2008).

Furthermore, although we found support for the effectiveness of moral reframing with regard to both attitudes toward the candidates and behavioral intentions, we did not use measures of real behavior (cf. Wolsko et al., 2016). Potentially tracking participants' actual voting behavior after exposure to reframed messages would be a promising route for future research. Furthermore, we only examined messages that appealed to the fairness and loyalty foundation. We chose these foundations as our examination of popular media suggested much of the information published about the candidates fit within a fairness or loyalty argument frame. Even so, it is an open question for future research how influential arguments couched in the other moral foundations might be.

Additionally, our two studies found an only partially consistent pattern in their moderated mediation analyses. In addition, the results of the second study should be considered with caution considering the nonsignificant interaction effects. Future research is needed to explore the mechanisms underlying the moral reframing effect in the political elections domain. Finally, in the present research, we did not have control conditions, and therefore it is impossible to know for sure which of the two conditions in the studies caused the persuasion effects we found. However, past research has used control conditions and found that the effect is in line with the moral reframing hypothesis (Feinberg & Willer, 2013, 2015), and as such, we feel confident that the effects we found were due to the morally reframed conditions.

Overall, our findings add to the growing body of research demonstrating how important it is to recognize and understand the moral values of those who take an opposing political position (Feinberg & Willer, 2013, 2015; Kidwell et al., 2013; Wolsko et al., 2016). As a whole, this literature highlights that the more individuals take the moral perspective of those who do not agree with them into consideration, the more successful they will be at reaching those individuals. The present research demonstrates that this is even the case in the context of one of the most politically polarizing events—political campaigns.

Supplemental Material

The supplemental material is available in the online version of the article.

References

Brandt, M. J., Reyna, C., Chambers, J. R., Crawford, J. T., & Wetherell, G. (2014). The ideological-conflict hypothesis intolerance among both liberals and conservatives. *Current Directions in Psychological Science*, 23, 27–34.

Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., & Barbaranelli, C. (2006). Personality and politics: Values, traits, and political choice. *Political Psychology*, 27, 1–28.

Chen, F. F., Jing, Y., & Lee, J. M. (2012). "I" value competence but "we" value social competence: The moderating role of voters' individualistic and collectivistic orientation in political elections. *Journal of Experimental Social Psychology*, 48, 1350–1355.

Cummings, J. (2008, November 5). 2008 campaign costliest in U.S. history. *Politico*. Retrieved from http://www.politico.com/story/2008/11/2008-campaign-costliest-in-us-history-015283

Day, M. V., Fiske, S. T., Downing, E. L., & Trail, T. E. (2014). Shifting liberal and conservative attitudes using moral foundations theory. *Personality and Social Psychology Bulletin*, 40, 1559–1573.

Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, 24, 56–62.

Feinberg, M., & Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, 41, 1665–1681.

Gallup.com. (n.d.). *Election polls—Presidential vote by groups*. Retrieved from http://www.gallup.com/poll/139880/election-polls-presidential-vote-groups.aspx

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106, 148–168.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029–1046.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101, 366–385.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98–116.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133, 55–66.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Janoff-Bulman, R., Sheikh, S., & Baldacci, K. G. (2008). Mapping moral motives: Approach, avoidance, and political orientation. *Journal of Experimental Social Psychology*, 44, 1091–1099.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129, 339–375.

Kidwell, B., Farmer, A., & Hardesty, D. M. (2013). Getting liberals and conservatives to go green: Political ideology and congruent appeals. *Journal of Consumer Research*, 40, 350–367.

Lakoff, G. (2002). *Moral politics: How liberals and conservatives think*. Chicago, IL: University of Chicago Press.

Morgan, G. S., Skitka, L. J., & Wisneski, D. C. (2010). Moral and religious convictions and intentions to vote in the 2008 presidential election. *Analyses of Social Issues and Public Policy*, 10, 307–320.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, 5, 359–371.

Skitka, L. J., & Bauman, C. W. (2008). Moral conviction and political engagement. *Political Psychology*, 29, 29–54.

Thorisdottir, H., Jost, J. T., Liviatan, I., & Shrout, P. E. (2007). Psychological needs and values underlying left-right political orientation: Cross-national evidence from Eastern and Western Europe. *Public Opinion Quarterly*, 71, 175–203.

Toner, K., Leary, M. R., Asher, M. W., & Jongman-Sereno, K. P. (2013). Feeling superior is a bipartisan issue extremity (not direction) of political views predicts perceived belief superiority. *Psychological Science*, 24, 2454–2462.

Trevino, L. K., Hartman, L. P., & Brown, M. (2000). Moral person and moral manager: How executives develop a reputation for ethical leadership. *California Management Review*, 42, 128–142.

Wolsko, C., Ariceaga, H., & Seiden, J. (2016). Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology*, 65, 7–19.

Author Biographies

Jan G. Voelkel is a research master student at Tilburg University. He is interested in moral and political psychology, with a focus on the question how to overcome the divide between liberals and conservatives.

Matthew Feinberg is an assistant professor of Organizational Behaviour at the Rotman School of Management. His research explores which psychological processes lead individuals to form cohesive groups and societies, with a focus on morality and political attitudes.

Handling Editor: Jesse Graham

ARTICLE

WILEY

Moral reframing: A technique for effective and persuasive communication across political divides

Matthew Feinberg[1] | Robb Willer[2]

[1]Rotman School of Management, University of Toronto

[2]Stanford University

Correspondence
Matthew Feinberg, Rotman School of Management, University of Toronto.
Email: matthew.feinberg@rotman.utoronto.ca

Abstract

The political landscape in the US and many other countries is characterized by policy impasses and animosity between rival political groups. Research finds that these divisions are fueled in part by disparate moral concerns and convictions that undermine communication and understanding between liberals and conservatives. This "moral empathy gap" is particularly evident in the moral underpinnings of the political arguments members of each side employ when trying to persuade one another. Both liberals and conservatives typically craft arguments based on their own moral convictions rather than the convictions of the people they target for persuasion. As a result, these moral arguments tend to be unpersuasive, even offensive, to their recipients. The technique of *moral reframing*—whereby a position an individual would not normally support is framed in a way that is consistent with that individual's moral values—can be an effective means for political communication and persuasion. Over the last decade, studies of moral reframing have shown its effectiveness across a wide range of polarized topics, including views of economic inequality, environmental protection, same-sex marriage, and major party candidates for the US presidency. In this article, we review the moral reframing literature, examining potential mediators and moderators of the effect, and discuss important questions that remain unanswered about this phenomenon.

1 | INTRODUCTION

> *"It is necessary to have regard to the person whom we wish to persuade ... what principles he acknowl-*
> *edges ... and then observe in the thing in question what affinity it has with the acknowledged principles."*
> --Blaise Pascal, *Pensees* (1669/1976)

Politics in the US and many other countries are characterized by high levels of polarization (e.g., Boxell, Gentzkow, & Shapiro, 2017; Groskopf, 2016; Pew Research Center, 2014; van der Eijk, Schmitt, & Binder, 2005), with liberals and conservatives disagreeing on political issues and also harboring deep animosities toward one another.[1] Research finds that both attitude polarization and antagonism between political groups are associated with underlying moral divisions (Haidt, 2012; Haidt & Graham, 2007; Koleva, Graham, Iyer, Ditto, & Haidt, 2012). Moral foundations theory (Graham, Haidt, & Nosek, 2009) describes the moral domain in terms of five distinct foundations that underlie moral attitudes and judgments: *care*—relating to the prevention and mitigation of suffering; *fairness*—relating to equality and discrimination concerns; *loyalty*—relating to the prioritization of one's group and its needs; *authority*—relating to showing respect for traditions and high ranking others; *sanctity*—relating to the protection of purity and sacredness (Graham et al., 2009; Graham et al., 2011).[2] Survey results find that, relative to conservatives, liberals more strongly endorse the care and fairness foundations, grounding many of their political positions in notions of compassion, nurturance, and social justice. Conservatives, in contrast, more strongly endorse the loyalty, authority, and sanctity foundations, and therefore ground many of their political stances in patriotism, traditionalism, and religious purity concerns (Caprara, Schwartz, Capanna, Vecchione, & Barbaranelli, 2006; Graham et al., 2009; Haidt & Graham, 2007; Jost, Glaser, Kruglanski, & Sulloway, 2003; Lakoff, 1996; but see also Voelkel & Brandt, 2019).

The rhetoric employed by advocates on the left and right often reflects these moral differences (Lakoff, 1996; Marietta, 2008),[3] with liberals arguing for policy positions in terms of the care and fairness foundations, and conservatives arguing for policy positions in terms of the loyalty, authority, and sanctity foundations (Feinberg & Willer, 2013, 2015). Use of moral rhetoric in this way may be an effective means for persuading, uniting, and motivating those who already share one's moral and political convictions, but it is likely to be an ineffective, even counterproductive, strategy for persuading political rivals. Indeed, research suggests that exposure to the other side's moral rhetoric can even result in increased commitment to one's existing stance and greater animosity toward those on the other side (Feinberg, Kovacheff, Teper, & Inbar, 2019; Gadarian & van der Vort, 2018; Skitka, Bauman, & Sargis, 2005; Tetlock, Kristel, Elson, Green, & Lerner, 2000), thereby exacerbating political polarization.

Recent research, however, suggests a means by which political advocates can persuade those who hold a different political ideology. *Moral reframing* is a technique in which a position an individual would not normally support is framed in a way that it is consistent with that individual's moral values. For example, an appeal in support of same-sex marriage is likely to garner greater support from conservatives if it is framed in terms of patriotism (e.g., "Gay Americans are proud and patriotic Americans") than if it were framed in terms of equality (e.g., "Gay Americans deserve equal rights"), because patriotism is a value more associated with conservatism, while equality is more associated with liberalism. Below, we summarize the emerging literature on moral reframing, highlight key demonstrations, and discuss when and why the technique is effective. We also point to some unanswered questions relating to moral reframing, puzzles that we hope will inspire future research on the topic.

1.1 | A review of moral reframing research

Framing is the use of a "central organizing idea or story line that provides meaning (to an issue or event)" (Gamson & Modigliani, 1987, p. 143; see also Druckman, 2001; Gitlin, 1980; Scheufele & Iyengar, 2012; Tversky & Kahneman, 1981). When the guiding force underlying the story line is grounded in moral concerns, this can be considered "moral framing" (e.g., Lakoff, 1996, 2004). Results of moral framing research across disciplines points to its effectiveness as a persuasion tool, especially in political discourse (e.g., Andrews, Clawson, Gramig, & Raymond, 2017; Barker, 2005;

Dehghani, Sagae, Sachdeva, & Gratch, 2014; Feinberg et al., 2019; Hoover, Johnson, Boghrati, Graham, & Dehghani, 2018; Lakoff, 2004; Marietta, 2008; Valenzuela, Piña, & Ramírez, 2017; van Zant & Moore, 2015).

Of note, however, persuasion research highlights the fundamental importance of matching an argument or frame to the message recipient's personal concerns and values (e.g., Boote, 1981; Hirsh, Kang, & Bodenhausen, 2012; Shen, 2004; Shen & Edwards, 2005; Watt, Maio, Haddock, & Johnson, 2008). Without such a match, the persuasive argument is less likely to resonate with the recipient and therefore will have little influence. Along these lines, the effectiveness of moral framing will be bounded by how closely the morality underlying the argument fits with the moral convictions of the target (Feinberg & Willer, 2013, 2015). Indeed, a moral frame that appeals to some individuals (e.g., liberals) could even offend others (e.g., conservatives), if the two groups ascribe to different moral foundations (Ditto & Koleva, 2011; Koleva et al., 2012). As a result, moral appeals that move one group, can actually backfire and result in the other group becoming more strident in its opposition. For instance, because conservatives strongly endorse the sanctity foundation, purity-based arguments such as those often made against immigration ("immigrants pollute society" Cisneros, 2008; "immigrants contaminate the purity of our nation" Valverde, 2008) may uniquely resonate with conservatives. However, because liberals do not endorse this foundation and often view it as contradictory to more foundations they do endorse (e.g., fairness), such arguments fail to persuade and may even backfire, strengthening liberals' pro-immigration stances (Gadarian & van der Vort, 2018; Kahan, Braman, Slovic, Gastil, & Cohen, 2007; Marietta, 2008; Tetlock et al., 2000).

A specific type of moral framing that can be effective for speaking across political divides is *moral reframing* (Feinberg & Willer, 2013, 2015; Voelkel & Feinberg, 2018). In the political arena, moral reframing involves arguing in favor of a political position that members of a political group would not normally support in terms of moral concerns that the members strongly ascribe to. Fitting a message to a particular audience in this way is persuasive because it makes the position relevant to and concordant with the audience's deeply held moral convictions. Seen through this lens, the purpose of moral reframing is to transform positions that would otherwise seem morally wrong to an audience, into something morally acceptable or even desirable.

Evidence for the effectiveness of moral reframing has built over the last decade. Most moral reframing research has focused on persuading liberals and conservatives to be more supportive of the other side's policy positions. In particular, a series of articles have examined how morally reframed arguments about the environment and climate change can persuade conservatives to be more supportive of pro-environmental policies and behavior (Feinberg & Willer, 2013; Feygina, Jost, & Goldsmith, 2010; Kidwell, Farmer, & Hardesty, 2013; McCright, Charters, Dentzman, & Dietz, 2016; Whitmarsh & Corner, 2017; Wolsko, Ariceaga, & Seiden, 2016). These studies have shown that conservatives are more persuaded by pro-environmental arguments that appeal to the loyalty, authority, and sanctity foundations. For example, Feygina et al. (2010) found that those more apt to be conservative (measured in their study in terms of the highly related trait of system justification) reported greater pro-environmental intentions and were more likely to sign a pro-environmental petition when presented with arguments emphasizing that being pro-environmental is consistent with the "American way of life" and therefore a patriotic duty. Similarly, Feinberg and Willer (2013) found that conservatives reported greater environmental concern, support for pro-environmental legislation, and belief in climate change when presented with a purity-based argument that emphasized how dirty, disgusting, and impure environmental degradation is, compared with a more typical, harm-focused argument that emphasized the devastation and dangers a failing environment can cause (for similar findings, see also Kidwell et al., 2013; Maibach et al., 2013; Wolsko et al., 2016).

Research shows that moral reframing is an effective means for persuasion on political issues far beyond the environment (Bloemraad, Silva, & Voss, 2016; Feinberg & Willer, 2015; Franks & Scherr, 2019; Simon & Gilliam, 2013; Thomas, 2017). For example, Feinberg and Willer (2015) found that liberals were more supportive of military spending when presented with fairness-focused arguments emphasizing the military's role in helping to overcome income inequality and racial discrimination than they were presented with more typical, authority- and loyalty-focused arguments that emphasized that the military demonstrated American patriotism and world dominance. In a representative sample of registered voters in California, Bloemraad et al. (2016) found that arguments in favor of legalizing the

status of undocumented immigrants were effective in persuading conservatives only when the argument appealed to the loyalty foundation, in particular to family unity. In a field experiment, Kalla, Levine, and Broockman (2019) found that door-to-door canvassers' pro-choice appeals were persuasive if articulated in terms of a moral value canvassers' elicited from residents early in the interaction.

A second category of moral reframing research has focused on voters' support for political candidates. This research builds on findings showing that individuals who more strongly endorse the care and fairness moral foundations tend to vote for Democratic candidates while individuals who strongly endorse the loyalty, authority, and sanctity foundations tend to vote for Republican candidates (Enke, 2018; Franks & Scherr, 2015). Voelkel and Feinberg (2018) presented American participants with arguments for why they should not support Donald Trump for president of the US. The arguments emphasized that Trump was unfit for office, either because he discriminates against minorities and promotes prejudice (fairness frame), or because he showed disloyalty to the country by dodging the draft during the Vietnam War (loyalty frame). Results showed that conservative participants were less supportive of Donald Trump's candidacy when they were presented with the loyalty frame. Additionally, a preregistered, nationally representative study found that an economically progressive candidate who articulated his platform in terms of more conservative values like patriotism and cultural traditionalism was supported more by self-identified conservatives, as well as moderates (Voelkel & Willer, 2019).

Finally, a third category of moral reframing research has explored its efficacy in shaping liberals' and conservatives' attitudes on issues that are largely nonpolitical or, at least, not strongly aligned with either liberal or conservative political ideologies. These studies have targeted outcomes like vaccine hesitancy, diversity and inclusion, and stem cell technology (Amin et al., 2017; Brannon, Carter, Murdock-Perriera, & Higginbotham, 2018; Clifford, Jerit, Rainey, & Motyl, 2015; Hoover et al., 2018; Kljajic & Feinberg, 2019; Lee, Yoon, Lee, & Royne, 2018; Winterich, Mittal, & Aquino, 2013). Along these lines, a handful of studies have examined the effects of moral reframing on liberals' and conservatives' tendencies to donate money to different types of charities. In a broad examination of moral reframing's impact on the donation behavior of liberals and conservatives, Kljajic and Feinberg (2019) found that certain charities were recognized as more "liberal charities" or "conservative charities," and morally reframing these charities' messaging led liberals to donate more to the "conservative charities" and conservatives to donate more to the "liberal charities." Similarly, Lee et al., (2018) examined the effects of different types of moral arguments in charity advertising on both liberals' and conservatives' donation behaviors, finding that arguments appealing to equality-based moral concerns were much more effective for liberals than conservatives (see also Winterich et al., 2013).[4]

2 | MECHANISMS DRIVING MORAL REFRAMING EFFECTS

Why does moral reframing work? The primary explanation is that morally reframed messages are influential because targets perceive a "match" between their moral convictions and the argument in favor of the other side's policy position. Moral convictions are one of the key underlying bases of people's attitudes. They are central and immutable parts of one's identity (Aquino & Reed, 2002; Atran & Axelrod, 2008; Kovacheff, Schwartz, Inbar, & Feinberg, 2018; Skitka et al., 2005; Strohminger & Nichols, 2014). Because moral convictions are so strongly held, arguments that appeal to them are difficult to discount, even when used to argue for a position one would typically oppose. As a result, when individuals face a morally reframed argument that resonates with their fundamental moral convictions, they are more likely to evaluate the argument positively and revise their relevant attitudes as a result. In line with this theorizing, moral reframing studies have found that perceived overlap between the message's argument and the moral convictions of the target mediates the effect (Feinberg & Willer, 2015).

This perceived overlap between a message's moral basis and the target's moral convictions may also give rise to subsequent mediating mechanisms. For instance, the overlap may evoke feelings of comfort or familiarity in the message recipient, making the argument easier to comprehend and process, encouraging a positive evaluation. Along these lines, Kidwell et al. (2013) found that conservatives' increase in pro-environmental intentions after reading

morally reframed environmental messages was mediated by the extent to which they reported that the message was clear and relatable (i.e., processing fluency; see also Gantman & Van Bavel, 2014). Additionally, moral reframing may signal to the message recipient that the message source is a fellow ingroup member and therefore someone who is credible and trustworthy (Cohen, 2003; Kalkhoff & Barnum, 2000). Wolsko et al., (2016), for example, found that conservatives presented with morally reframed environmental appeals were persuaded largely because they perceived the values being expressed in the message as coming from a conservative source.

2.1 | Potential moderators of moral reframing effects

Although a number of studies have shown the efficacy of moral reframing, potential moderating factors that either strengthen or weaken the effect remain largely unexplored. One category of potential moderators concerns features of the message itself. For instance, a morally reframed message that fails to connect a policy position to the moral foundations of the target audience will likely be ineffective. This might occur because the association with that moral foundation is not emphasized strongly enough to resonate with the target, because the argument made does not make a compelling case for viewing the policy as promoting the moral foundation, or possibly because the moral argument is obscured by other nonmoral arguments also included in the message. These possible moderating factors may help explain some studies that have found inconsistent evidence for moral reframing (e.g., Day, Fiske, Downing, & Trail, 2014; Peterson & Simonovits, 2017; van de Rijt, Akin, Willer, & Feinberg, 2016).

Conversely, a morally reframed argument might be particularly effective if it evokes strong moral emotions related to the moral foundation being targeted. Research has shown that different moral emotions correspond with, and reinforce, different foundations (Horberg, Oveis, & Keltner, 2011). For instance, the sanctity foundation corresponds closely with the emotion of disgust (Horberg, Oveis, Keltner, & Cohen, 2009; Wisneski & Skitka, 2017). With this in mind, morally reframed messages targeting conservatives via the sanctity foundation should be particularly potent if the message elicits moral disgust in the target. Consistent with this reasoning, Feinberg and Willer (2013) found that the persuasive effect of a purity-based argument in support of environmentalism (e.g., "pollution in our environment inevitably contaminates us and our bodies") among conservatives was mediated by conservatives' self-reported disgust after reading the message. Although this research focused on disgust as a mediator rather than a moderator, these results suggest that a sanctity message that elicits disgust may be more effective at persuading conservatives than one that does not.

The effectiveness of a morally reframed message may also depend on which aspects of the targeted moral foundation the message highlights. Although the relationship between political ideology and each of the five moral foundations has proven robust (Graham et al., 2011), certain aspects of the different foundations are likely more fundamental to liberal and conservative morality. For instance, although liberals very strongly endorse the care foundation, this may be primarily due to liberals' emphasis on compassion and alleviating the suffering of those most in need, rather than other aspects relating to protection and security concerns (Hirsh, DeYoung, Xu, & Peterson, 2010; Piurko, Schwartz, & Davidov, 2011). As such, morally reframed messages that speak to concerns about being compassionate and alleviating suffering would likely be more persuasive for liberals than emphasizing national security (Feinberg & Willer, 2015). Similarly, egalitarianism and social justice may be aspects of the fairness foundation that resonate most strongly with liberals, compared to other aspects of fairness like proportionality (i.e., equity; Haidt, 2012; Jost et al., 2003). As a result, morally reframed messages that speak to these values are likely to be more persuasive for liberals. Indeed, in one moral reframing experiment, liberals were more likely to donate to a charity when the message appealed to egalitarian concerns, but when the argument appealed to proportionality fairness concerns, conservatives were more persuaded (Lee et al., 2018). Conversely, for conservatives, aspects of the loyalty foundation pertaining to national pride and patriotism would likely be more effective in moral reframing than loyalty appeals beyond the national level (see Feinberg & Willer, 2015; Feygina et al., 2010; Wolsko et al., 2016).

Message source will likely also have a strong moderating effect on the impact of moral reframing. For example, in a study of moral messaging around environmental issues, Wolsko et al., (2016) found that one reason moral reframing

worked was because message recipients perceived the source to be part of their political ingroup. But what would happen if the message source was a known political outgroup member? One recent study (Voelkel & Willer, 2019) found that conservatives were significantly more supportive of a Democratic presidential candidate if the candidate appealed to conservative value concerns such as patriotism and respect for tradition than if he appealed to more liberal value concerns like social justice and egalitarianism. This suggests that moral reframing effects can be effective enough to be persuasive, even when seen as coming from a political outgroup. At the same time, it is unlikely that moral reframing would be effective if the source was a reviled political outgroup member. For example, an argument in favor of military spending that highlights how the military provides a level playing field for the poor and minorities in the US would likely be unpersuasive to liberals if the messenger was Sean Hannity, Rush Limbaugh, or Donald Trump. Political psychology research suggests that the effect of moral reframing would be substantially muted, if not eliminated, in this situation (Cohen, 2003; see also Goren, 2005), because the target would approach the message with outgroup bias and would be strongly motivated to reject the argument before even reading it (Kahan, 2015).

Lastly, national context is likely another important moderating factor. Although in many ways, political ideology manifests similarly across cultural and national contexts (Lakoff, 1996), the bases of economic and social ideology can vary markedly across nations and political systems (Malka, Soto, Inzlicht, & Lelkes, 2014). With this in mind, a moral reframing strategy for economic policy in some countries might need to be quite different than it would be in the US. One example might be post-Soviet countries, where conservatives often support greater government regulation of the economy because of the historical legacy of communism as the traditional status quo. In these countries, arguments for greater regulation of the economy would need to target liberals, linking government regulation with the more liberal moral foundations to be effective.

2.2 | Effects of moral reframing on untargeted individuals

It is worth noting that most moral reframing studies to date have found nonsignificant effects for the untargeted group—that is, the group that already supports the policy stance being argued for. For example, Feinberg and Willer (2015) found that liberals showed similar levels of support for universal health care regardless of whether they read a typical moral message appealing to equality or a reframed message appealing to purity concerns. Likewise, conservatives showed similar levels of support for increased military spending regardless of whether they read a message appealing to loyalty or egalitarian concerns. These null effects suggest that untargeted groups were not persuaded by messages grounded in moral language they oppose, nor by messages grounded in moral language they endorse. Bolstering this point, in the handful of moral reframing studies with a control condition, the untargeted group's level of support did not differ across any condition. It appears that in situations where the argument is congruent with one's policy stance, how that argument is made may matter little.

These null results for the untargeted group have important implications for individuals speaking to audiences consisting of both liberals and conservatives. If the type of messaging has no effect on the untargeted group, but morally reframed messages positively influence the targeted group, then speakers in these situations should rely on morally reframed messages. However, we believe future research should explore this in more depth. Existing moral reframing research may have found the null effects because it has focused primarily on highly polarized political issues where the untargeted group's level of support is close to a ceiling, leaving little room for increases in support. Nonetheless, in situations where the untargeted group's support is not crystallized, moral messaging could still have an impact. Additionally, even if moral messages do not affect the untargeted group attitudinally, they may still have behavioral effects. Considering moral motivations are important for collective action (van Zomeren, Postmes, Spears, & Bettache, 2011), moral messages that appeal to existing supporters may not strengthen their support, but still influence their willingness to act (Marietta, 2008, however, see Voelkel & Willer, 2019). With this in mind, it may behoove individuals speaking to diverse audiences to employ a mix of moral frames—some appealing to existing supporters, others appealing to nonsupporters. We hope future research might explore the benefits (or possible pitfalls) of using such mixed frames.

2.3 | Why people do not use moral reframing

Even though research suggests moral reframing is an effective strategy for persuading those on the other side of political divides, people rarely employ the strategy spontaneously. Feinberg and Willer (2015) found that fewer than 10% of liberal or conservative participants wrote morally reframed appeals when asked to write political arguments aimed at persuading their political counterparts and despite being offered a cash prize for effectively doing so. If moral reframing is typically effective, why is it so rarely employed?

One possibility is that people do not realize that moral reframing is effective. Research indicates that individuals experience their moral convictions as objective truths about the world (Skitka et al., 2005). As a result, it can be difficult to recognize that there are different "truths" that other people believe in (Ditto & Koleva, 2011; Kovacheff et al., 2018). Indeed, polling data indicates that people are apt to perceive someone who does not endorse their morality as simply immoral or evil, rather than morally different (Doherty & Kiley, 2016). Without recognizing that one's political rivals possess different morals, and without a clear understanding of what those different morals are, using moral reframing becomes impossible. Moreover, even if individuals are aware of the moral differences, they may still not use moral reframing because a morally reframed argument might intuitively seem weak or ineffectual because it does not resonate with them personally.

Alternatively, individuals may have the knowledge needed to craft morally reframed messages, but refuse to do so on principle. Strategically making arguments that appeal to moral values that one does not strongly endorse may feel unethical. In fact, research on the concept of "taboo trade-offs" (Tetlock, 2003) finds that even entertaining the possibility of compromising one's moral convictions to achieve strategic ends—such as buying body parts for transplants or hiring another person to serve one's prison sentence—can feel immoral and unacceptable. In the context of moral foundations, for many liberals, aspects of the authority foundation may contradict moral beliefs about equality and progressivism. As a result, making arguments grounded in the authority foundation could feel immoral. Moreover, individuals might also fear that others, particularly those who share their moral beliefs, will judge them harshly for appealing to "false morals". Finally, people might be averse to making morally reframed arguments because they believe doing so could indirectly reinforce values they disagree with (Lakoff, 2004), perhaps leading to short run gains on a given issue, but serving to validate moral values that are anathema to their moral and political worldview in the long-term.

In initial explorations of why most individuals do not use moral reframing, Feinberg and Willer (2015) found that when participants were asked to identify whether a morally reframed argument or a more typical moral argument would be more persuasive to those on the other side of the political spectrum, 64% of liberal participants and 85% of conservative participants correctly selected the morally reframed argument. This suggests that most individuals can recognize the persuasive appeal of morally reframed arguments even if they do not spontaneously use them. In addition, results indicated that when participants were asked to choose one of the arguments to send to a member of the rival ideological group they were tasked with persuading, 80% of liberals and 94% of conservatives who previously indicated the morally reframed argument would be more persuasive also chose to send the morally reframed argument, suggesting that most individuals would not object to using morally reframed arguments on principle. Taken together, these findings suggest that the most likely barriers to individuals using moral reframing spontaneously are (a) they do not think to use it, or (b) they do not know how to, rather than an inability to recognize the persuasive power of these arguments, or an unwillingness to use them. However, more research is needed to fully understand why few use morally reframed arguments.

3 | CONCLUSION

In his classic of applied social psychology, *How to Win Friends and Influence People*, Dale Carnegie (2017 [1936]) underscored the importance of accounting for the interests and perspective of a target of persuasion:

For example, one day Ralph Waldo Emerson and his son tried to get a calf into the barn. But they made the common mistake of thinking only of what they wanted; so he stiffened his legs and stubbornly refused to leave the pasture. The Irish housemaid saw their predicament. She couldn't write essays and books; but on this occasion at least, she had more horse sense, or calf sense, than Emerson had. She thought of what the calf wanted; so she put her maternal finger in the calf's mouth and let the calf suck her finger as she gently led him into the barn.

This principle applies not only to the finger suckling of recalcitrant calves, but also—as we find in the above-reviewed studies—to bridging one of the most fundamental divides in American society. A growing roster of studies attest to the potential effectiveness of moral reframing as a persuasive tool.

In all, moral reframing offers a means for creating political consensus, which can have many benefits. In deeply divided polities, such as the contemporary US, consensus may be a valuable end to itself, facilitating effective governance and preserving the legitimacy of democratic institutions. For advocates, moral reframing may be attractive as an alternative to simple compromise on political issues, at present the dominant model for building political consensus. Indeed, one prior study found that subjectively reframing the apparent moral virtues of a policy was more impactful than the extremity of that policy on a liberal–conservative spectrum, suggesting compromise might often be less effective than moral reframing (e.g., Voelkel & Willer, 2019).

Further, beyond being effective for building consensus, moral reframing may produce other valuable outcomes. Moral reframing may reduce animosity individuals feel for those on the other side of a significant political divide (Doherty & Kiley, 2016). Though people would likely use moral reframing for the strategic purpose of winning over their rivals, effectively using the technique requires developing a clear understanding of what the other side values and believes. Thus, inherent in using moral reframing is seeking out and understanding the moral perspective of one's rivals. Developing a more accurate understanding of others' perspectives may also have downstream effects on affective polarization, which research shows stems in part from exaggerated and inaccurate stereotypes of political outgroup members (Farwell & Weiner, 2000; Graham, Nosek, & Haidt, 2012). When individuals have a better sense of a rival group's perspective, they will be less likely to degrade them or view them as clearly evil. Future research should explore the possibility that moral reframing of divisive political issues could be helpful in this way to reduce antipathy and negative affect between rival political groups.

Of course, it would be naïve to suggest moral reframing will always have positive effects. In fact, many of the greatest atrocities in human history have been supported by political projects that employed strategic moral reframing. Where the Third Reich employed purity, loyalty, and authority arguments, Stalin and the Khmer Rouge built their regimes on the values of authority and fairness. These and countless other examples show that, like any pragmatic tool, moral reframing can be no more moral than the ends to which it is put. That said, while any technique of political persuasion could in principle be used to ameliorate intractable and destructive political divides, or to justify cruelty and brutality, there is reason to hope that the everyday practice of moral reframing, because its effective use requires careful attention to the perspective of one's political rivals, could foster that rarest of things in a bitterly divided society—mutual understanding.

ENDNOTES

[1] Whereas in the US, the political divide is between "liberals" and "conservatives" (or Democrats and Republicans), both the substance of political divides and the terms used to describe them vary across cultural contexts (Malka et al., 2014). However, research suggests that the liberal–conservative divide on social issues in particular manifests in similar ways across cultures (e.g., Feinberg, Wehling, Chung, Saslow, & Melvær Paulin, 2019; Graham et al., 2011).

[2] There is also evidence of an additional moral foundation relating to *freedom* (see Iyer, Koleva, Graham, Ditto, & Haidt, 2012).

[3] Although we focus primarily on liberals versus conservatives in our review, the reasoning and insights from this review apply as well to the highly related political divisions between the left and right and Democrats and Republicans.

[4] In addition, they found that arguments appealing to proportionality-based rewards were more effective for influencing conservatives' donation behavior. These findings speak to the possibility that the fairness moral foundation may have two components—one that relates to egalitarian concerns and one that relates to proportionality concerns, with the former being a central moral concern of liberals, and the latter being of concern to conservatives (Haidt, 2012).

ORCID

Matthew Feinberg https://orcid.org/0000-0001-8626-1620

REFERENCES

Amin, A. B., Bednarczyk, R. A., Ray, C. E., Melchiori, K. J., Graham, J., Huntsinger, J. R., & Omer, S. B. (2017). Association of moral values with vaccine hesitancy. *Nature Human Behaviour*, *1*(12), 873–880. https://doi.org/10.1038/s41562-017-0256-5

Andrews, A. C., Clawson, R. A., Gramig, B. M., & Raymond, L. (2017). Finding the right value: Framing effects on domain experts. *Political Psychology*, *38*(2), 261–278. https://doi.org/10.1111/pops.12339

Aquino, K., & Reed, I. I. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423–1440. https://doi.org/10.1037/0022-3514.83.6.1423

Atran, S., & Axelrod, R. (2008). Reframing sacred values. *Negotiation Journal*, *24*(3), 221–246. https://doi.org/10.1111/j.1571-9979.2008.00182.x

Barker, D. (2005). Values, frames, and persuasion in presidential nomination campaigns. *Political Behavior*, *27*(4), 375–395. https://doi.org/10.1007/s11109-005-8145-4

Bloemraad, I., Silva, F., & Voss, K. (2016). Rights, economics, or family? Frame resonance, political ideology, and the immigrant rights movement. *Social Forces*, *94*(4), 1647–1674. https://doi.org/10.1093/sf/sov123

Boote, A. S. (1981). Market segmentation by personal values and salient product attributes. *Journal of Advertising Research*, *21*, 29–35.

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences*, *114*(40), 10612–10617. https://doi.org/10.1073/pnas.1706588114

Brannon, T. N., Carter, E. R., Murdock-Perriera, L. A., & Higginbotham, G. D. (2018). From backlash to inclusion for all: Instituting diversity efforts to maximize benefits across group lines. *Social Issues and Policy Review*, *12*(1), 57–90. https://doi.org/10.1111/sipr.12040

Caprara, G. V., Schwartz, S., Capanna, C., Vecchione, M., & Barbaranelli, C. (2006). Personality and politics: Values, traits, and political choice. *Political Psychology*, *27*(1), 1–28. https://doi.org/10.1111/j.1467-9221.2006.00447.x

Carnegie, D. (2017). *How to win friends & influence people*. Förlag: Musaicum Books.

Cisneros, J. D. (2008). Contaminated communities: The metaphor of "immigrant as pollutant" in media representations of immigration. *Rhetoric & Public Affairs*, *11*(4), 569–601. https://doi.org/10.1353/rap.0.0068

Clifford, S., Jerit, J., Rainey, C., & Motyl, M. (2015). Moral concerns and policy attitudes: Investigating the influence of elite rhetoric. *Political Communication*, *32*(2), 229–248. https://doi.org/10.1080/10584609.2014.944320

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808–822. https://doi.org/10.1037/0022-3514.85.5.808

Day, M. V., Fiske, S. T., Downing, E. L., & Trail, T. E. (2014). Shifting liberal and conservative attitudes using moral foundations theory. *Personality and Social Psychology Bulletin*, *40*(12), 1559–1573. https://doi.org/10.1177/0146167214551152

Dehghani, M., Sagae, K., Sachdeva, S., & Gratch, J. (2014). Analyzing political rhetoric in conservative and liberal weblogs related to the construction of the "Ground Zero Mosque". *Journal of Information Technology & Politics*, *11*(1), 1–14. https://doi.org/10.1080/19331681.2013.826613

Ditto, P. H., & Koleva, S. P. (2011). Moral empathy gaps and the American culture war. *Emotion Review*, *3*(3), 331–332. https://doi.org/10.1177/1754073911402393

Doherty, C., & Kiley, J. (2016, June 22). Key facts about partisanship and political animosity in America. Retrieved from http://www.pewresearch.org/fact-tank/2016/06/22/key-facts-partisanship/.

Druckman, J. N. (2001). The implications of framing effects for citizen competence. *Political Behavior*, *23*, 225–256. https://doi.org/10.1023/A:1015006907312

Enke, B. (2018). *Moral values and voting* (No. w24268). National Bureau of Economic Research.

Farwell, L., & Weiner, B. (2000). Bleeding hearts and the heartless: Popular perceptions of liberal and conservative ideologies. *Personality and Social Psychology Bulletin*, *26*(7), 845–852. https://doi.org/10.1177/0146167200269009

Feinberg, M., Kovacheff, C., Teper, R., & Inbar, Y. (2019). Understanding the process of moralization: How eating meat becomes a moral issue. *Journal of Personality and Social Psychology*, *117*, 50–72. https://doi.org/10.1037/pspa0000149

Feinberg, M., & Willer, R. (2013). The moral roots of environmental attitudes. *Psychological Science*, *24*(1), 56–62. https://doi.org/10.1177/0956797612449177

Feinberg, M., & Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, *41*(12), 1665–1681. https://doi.org/10.1177/0146167215607842

Feinberg, M., Wehling, E., Chung, J. M., Saslow, L. R., & Melvær Paulin, I. (2019). Measuring Moral Politics: How Strict and Nurturant Family Values Explain Individual Differences in Conservatism, Liberalism, and the Political Middle. *Journal of Personality and Social Psychology*. Advance online publication.http://dx.doi.org/10.1037/pspp0000255

Feygina, I., Jost, J. T., & Goldsmith, R. E. (2010). System justification, the denial of global warming, and the possibility of "system-sanctioned change". *Personality and Social Psychology Bulletin*, *36*(3), 326–338. https://doi.org/10.1177/0146167209351435

Franks, A. S., & Scherr, K. C. (2015). Using moral foundations to predict voting behavior: Regression models from the 2012 US presidential election. *Analyses of Social Issues and Public Policy*, *15*(1), 213–232. https://doi.org/10.1111/asap.12074

Franks, A. S., & Scherr, K. C. (2019). Economic issues are moral issues: The moral underpinnings of the desire to reduce wealth inequality. *Social Psychological and Personality Science*, *10*(4), 553–562.

Gadarian, S. K., & van der Vort, E. (2018). The gag reflex: Disgust rhetoric and gay rights in American politics. *Political Behavior*, *40*(2), 521–543. https://doi.org/10.1007/s11109-017-9412-x

Gamson, W. A., & Modigliani, A. (1987). The changing culture of affirmative action. *Research in Political Sociology*, *3*, 137–177.

Gantman, A. P., & Van Bavel, J. J. (2014). The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. *Cognition*, *132*, 22–29. https://doi.org/10.1016/j.cognition.2014.02.007

Gitlin, T. (1980). *The whole world is watching: Mass media in the making & unmaking of the new left*. Berkeley: University of California Press.

Goren, P. (2005). Party identification and core political values. *American Journal of Political Science*, *49*(4), 881–896. https://doi.org/10.1111/j.1540-5907.2005.00161.x

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Nosek, B. A., & Haidt, J. (2012). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PLoS ONE*, *7*(12), e50092. https://doi.org/10.1371/journal.pone.0050092

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

Groskopf, C. (2016). European politics is more polarized than ever, and these numbers prove it. Retrieved from https://qz.com/645649/european-politics-is-more-polarized-than-ever-and-these-numbers-prove-it/

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Vintage.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Hirsh, J. B., DeYoung, C. G., Xu, X., & Peterson, J. B. (2010). Compassionate liberals and polite conservatives: Associations of agreeableness with political ideology and moral values. *Personality and Social Psychology Bulletin*, *36*(5), 655–664. https://doi.org/10.1177/0146167210366854

Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological Science*, *23*(6), 578–581. https://doi.org/10.1177/0956797611436349

Hoover, J., Johnson, K., Boghrati, R., Graham, J., & Dehghani, M. (2018). Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology*, *4*(1), 9. https://doi.org/10.1525/collabra.129

Horberg, E. J., Oveis, C., & Keltner, D. (2011). Emotions as moral amplifiers: An appraisal tendency approach to the influences of distinct emotions upon moral judgment. *Emotion Review*, *3*(3), 237–244. https://doi.org/10.1177/1754073911402384

Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology*, *97*(6), 963–976. https://doi.org/10.1037/a0017423

Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, *7*(8), e42366. https://doi.org/10.1371/journal.pone.0042366

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Kahan, D. M. (2015). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, 1-16.

Kahan, D. M., Braman, D., Slovic, P., Gastil, J., & Cohen, G. L. (2007). The second national risk and culture study: Making sense of-and making progress in-the American culture war of fact. *GWU Legal Studies Research Paper*, *370*, 8–26.

Kalkhoff, W., & Barnum, C. (2000). The effects of status-organizing and social identity processes on patterns of social influence. *Social Psychology Quarterly*, *63*, 95–115. https://doi.org/10.2307/2695886

Kalla, J. L., Levine, A. S., & Broockman, D. E. (2019). Personalizing moral reframing in interpersonal conversation: A field experiment. *Working paper*.

Kidwell, B., Farmer, A., & Hardesty, D. M. (2013). Getting liberals and conservatives to go green: Political ideology and congruent appeals. *Journal of Consumer Research*, *40*(2), 350–367. https://doi.org/10.1086/670610

Kljajic, M. & Feinberg, M. (2019). Moral foundations of charitable giving. *Working paper*.

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality*, *46*(2), 184–194. https://doi.org/10.1016/j.jrp.2012.01.006

Kovacheff, C., Schwartz, S., Inbar, Y., & Feinberg, M. (2018). The problem with morality: Impeding progress and increasing divides. *Social Issues and Policy Review*, *12*(1), 218–257. https://doi.org/10.1111/sipr.12045

Lakoff, G. (1996). *Moral politics: What conservatives know that liberals don't*. Chicago: University of Chicago Press.

Lakoff, G. (2004). *Don't Think of an Elephant! Know Your Values and Frame the Debate*. White River Junction, Vermont: Chelsea Green Publishing.

Lee, Y., Yoon, S., Lee, Y. W., & Royne, M. B. (2018). How liberals and conservatives respond to equality-based and proportionality-based rewards in charity advertising. *Journal of Public Policy & Marketing*, *37*(1), 108–118.

Maibach, E., Roser-Renouf, C., Vraga, E., Bloodhart, B., Anderson, A., Stenhouse, N., & Leiserowitz, A. (2013). A national survey of Republicans and Republican-leaning Independents on energy and climate change. *George Mason University Center for Climate Change Communication and Yale Project on Climate Change Communication*.

Malka, A., Soto, C. J., Inzlicht, M., & Lelkes, Y. (2014). Do needs for security and certainty predict cultural and economic conservatism? A cross-national analysis. *Journal of Personality and Social Psychology*, *106*(6), 1031–1051. https://doi.org/10.1037/a0036170

Marietta, M. (2008). From my cold, dead hands: Democratic consequences of sacred rhetoric. *The Journal of Politics*, *70*(3), 767–779. https://doi.org/10.1017/S0022381608080742

McCright, A. M., Charters, M., Dentzman, K., & Dietz, T. (2016). Examining the effectiveness of climate change frames in the face of a climate change denial counter-frame. *Topics in Cognitive Science*, *8*(1), 76–97. https://doi.org/10.1111/tops.12171

Pascal, B. (1669/1976). *Thoughts*. Paris: GF Flammarion.

Peterson, E., & Simonovits, G. (2017). Costly values: The limited benefits and potential costs of targeted policy justifications. *Journal of Experimental Political Science*, *4*(2), 95–106. https://doi.org/10.1017/XPS.2017.3

Pew Research Center. (2014). Political polarization in the American public. Retrieved from http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/

Piurko, Y., Schwartz, S. H., & Davidov, E. (2011). Basic personal values and the meaning of left-right political orientations in 20 countries. *Political Psychology*, *32*(4), 537–561. https://doi.org/10.1111/j.1467-9221.2011.00828.x

Scheufele, D. A., & Iyengar, S. (2012). The state of framing research: A call for new directions. In *The Oxford Handbook of Political Communication Theories* (pp. 1–26). New York: Oxford University Press.

Shen, F. (2004). Chronic accessibility and individual cognitions: Examining the effects of message frames in political advertisements. *Journal of Communication*, *54*(1), 123–137. https://doi.org/10.1111/j.1460-2466.2004.tb02617.x

Shen, F., & Edwards, H. H. (2005). Economic individualism, humanitarianism, and welfare reform: A value-based account of framing effects. *Journal of Communication*, *55*(4), 795–809. https://doi.org/10.1111/j.1460-2466.2005.tb03023.x

Simon, A. F., & Gilliam Jr., F. D. (2013). *Framing and facts*: *Necessary synergies in communicating about public safety and criminal justice*. Retrieved from https://www.frameworksinstitute.org/assets/files/pscj_values_and_facts.pdf

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology*, *88*(6), 895–917. https://doi.org/10.1037/0022-3514.88.6.895

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171. https://doi.org/10.1016/j.cognition.2013.12.005

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*(7), 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Thomas, C. (2017, June). *Universal Basic Income: The Meaning of Money Matters*. Presentation at the Society for the Psychological Study of Social Issues 2017 Conference, Albuquerque, NM. Presentation.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Valenzuela, S., Piña, M., & Ramírez, J. (2017). Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. *Journal of Communication*, *67*(5), 803–826.

Valverde, M. (2008). Racial purity, sexual purity, and immigration policy. In *The history of immigration and racism in Canada* (pp. 175–188). Toronto, ON: Canadian Scholars' Press.

Van De Rijt, A., Akin, I. A., Willer, R., & Feinberg, M. (2016). Success-breeds-success in collective political behavior: Evidence from a field experiment. *Sociological Science*, *3*, 940–950. https://doi.org/10.15195/v3.a41

Van der Eijk, C., Schmitt, H., & Binder, T. (2005). Left-right orientation and party choice. In J. S. Thomassen (Ed.), *The European voter. A comparative study of modern democracies* (pp. 167–191). Oxford: Oxford University Press. https://doi.org/10.1093/0199273219.003.0007

Van Zant, A. B., & Moore, D. A. (2015). Leaders' use of moral justifications increases policy support. *Psychological Science*, *26*(6), 934–943. https://doi.org/10.1177/0956797615572909

Van Zomeren, M., Postmes, T., Spears, R., & Bettache, K. (2011). Can moral convictions motivate the advantaged to challenge social inequality? Extending the social identity model of collective action. *Group Processes & Intergroup Relations*, *14*(5), 735–753. https://doi.org/10.1177/1368430210395637

Voelkel, J. G., & Brandt, M. J. (2019). The effect of ideological identification on the endorsement of moral values depends on the target group. *Personality and Social Psychology Bulletin*, *45*(6), 851–863.

Voelkel, J. G., & Feinberg, M. (2018). Morally reframed arguments can affect support for political candidates. *Social Psychological and Personality Science*, *9*(8), 917–924. https://doi.org/10.1177/1948550617729408

Voelkel, J. G., & Willer, R. (2019). Resolving the progressive paradox: Conservative value framing of progressive economic policies increases candidate support. *SSRN working paper*.

Watt, S. E., Maio, G. R., Haddock, G., & Johnson, B. T. (2008). Attitude functions in persuasion: Matching, involvement, self-affirmation, and hierarchy. In R. Prislin & W. Crano (Eds.), *Attitudes and Attitude Change* (pp. 189–211). Frontiers of Social Psychology Series. New York: Psychology Press.

Whitmarsh, L., & Corner, A. (2017). Tools for a new climate conversation: A mixed-methods study of language for public engagement across the political spectrum. *Global Environmental Change*, *42*, 122–135. https://doi.org/10.1016/j.gloenvcha.2016.12.008

Winterich, K. P., Mittal, V., & Aquino, K. (2013). When does recognition increase charitable behavior? Toward a moral identity-based model. *Journal of Marketing*, *77*(3), 121–134. https://doi.org/10.1509/jm.11.0477

Wisneski, D. C., & Skitka, L. J. (2017). Moralization through moral shock: Exploring emotional antecedents to moral conviction. *Personality and Social Psychology Bulletin*, *43*(2), 139–150. https://doi.org/10.1177/0146167216676479

Wolsko, C., Ariceaga, H., & Seiden, J. (2016). Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology*, *65*, 7–19.

AUTHOR BIOGRAPHIES

Matthew Feinberg is an Assistant Professor of Organizational Behavior and Human Resource Management at the Rotman School of Management at the University of Toronto. He earned his PhD in social-personality psychology from the University of California, Berkeley.

Robb Willer is a Professor of Sociology, Psychology (by courtesy), and Organizational Behavior (by courtesy) at Stanford University. He earned his PhD in sociology from Cornell University.

Implicit Bias and the Illusion of Mediator Neutrality

Carol Izumi[*]

INTRODUCTION

Plaintiff (P), the owner/operator of a carpet cleaning business, sued the defendant-homeowners for $500 in a breach of contract action for the unpaid balance of a $1,000 carpet cleaning agreement. Defendants (Ds or Mr. and Mrs. D) counterclaimed for the return of the $500 deposit they paid before work began. Ds hired P to dry out and clean the soaked carpet in their basement, which had flooded during a storm. Ds refused to pay the balance because the carpet had not dried out as P promised. Under the small claims court mediation program, the parties were required to attempt mediation before a trial date was set.

P was a middle-aged white male who attended the mediation in work clothes. Ds were an equally mature married couple of Asian descent who spoke with noticeable accents. They were dressed in what might be called "business casual" attire. The mediation was conducted around a large conference table by two white co-mediators: a male who looked to be in his forties and a younger female. The mediators conducted a "caucus model" facilitative-style mediation. P presented the case as a simple breach of contract: the agreement between the parties required the homeowners to make two

$500 payments and the second payment had not been made. Mr. D complained that the business owner was trying to cheat him by charging him for work that was unsatisfactory. During the mediation, P and Mr. D had markedly different demeanors. P was matter-of-fact and even-tempered. Mr. D was angry and agitated. Mrs. D sat quietly behind and to the right of her husband during the mediation. She spoke once and was quickly shushed by her husband.

In the joint session, P described the business transaction and his actions placing large fans in the basement to dry out the carpet. He stated that he had stressed to the homeowners the importance of keeping the upstairs door to the basement open for air to circulate. However, when he went to the house the following day, he found the door shut. P argued that the carpet did not dry as he expected because Ds did not keep the door open as instructed. The mediators asked P a number of questions about the contract, his interaction with Ds, and his professional cleaning techniques. When it was his turn to speak, Mr. D argued that P failed to complete the work as promised and that P's work was unsatisfactory. He asserted that the door was kept open as instructed; P saw it closed because Ds were preparing food and had temporarily shut the basement door in the kitchen because of the musty odor downstairs. During their co-mediator caucus after the joint session, the mediators commented that Ds failed to keep the door open.

In the individual sessions with the disputants, the mediators gathered and clarified information and explored options. P reiterated his position that he was entitled to the contract price since Ds' failure to keep the door open protracted the carpet drying process. In their individual session, Ds pressed that they were not satisfied with P's work because the carpet did not dry out in the promised time frame. Mr. D said he entered into the transaction cautiously because he was aware that American businesses sometimes take advantage of customers. After these two individual sessions, the mediators caucused and decided that the parties had reached an impasse. They brought the parties back together, conducted a bit more discussion, and concluded the session. The mediation was terminated in less than an hour without an agreement, and the matter was scheduled for trial. With more cases awaiting mediation, the mediators were quickly assigned another small claims case.

The preceding description is based on a small claims case mediation that I witnessed as a requirement for civil mediator certification in Michigan.[1] As an observer, I wondered why the mediator team decided that Ds failed to keep the door open despite their consistent assertions to the contrary. What judgments did the mediators make to reach such a determination? I was curious as to why the mediators failed to explore the door open/door closed issue in the individual sessions with the parties since it seemed significant. What factors and phenomena might have influenced the mediators' thought processes, judgment, and decision-making? I immediately thought about the possibility that racial dynamics played a role. None of the other observers I asked imagined that racial issues were at play. Being the sole non-white observer, perhaps I was more sensitive to potential racial aspects in the mediation.

One could view this mediation in a number of ways. When I presented this scenario to a group of mediation academics, one colleague opined that it was simply an example of bad mediation. In his view, the mediators seemed poorly skilled and their process lacked a systematic exploration of party interests, goals, priorities, and options. To him, the mediators were guilty of incompetence, nothing more. Another colleague supposed that the mediators were pressured by time limits and a waiting room full of parties in other cases. To this colleague, it was merely an example of "speed mediation." A third professor reasoned that the mediators made a credibility determination and decided that P was more believable. She allowed that mediators make credibility calls all the time and acknowledged that race could play a role in determining credibility. For all three mediation experts, nothing in the scenario raised concerns about mediator neutrality. I offer this mediation scenario as an opportunity to explore the nuances of mediator neutrality, consider the pervasiveness of unconscious bias, and provoke new dialogue.

This Article probes the complex challenges of a mediator's ethical duty to mediate disputes in a neutral manner against the behavioral

---

1. Observation of two mediations was required as part of the Michigan civil mediator certification process. MEDIATOR TRAINING STANDARDS AND PROCEDURES § 5.2.3 (Office of Dispute Resolution, Mich. Supreme Court 2005), *available at* http://courts.michigan.gov/scao/resources/standards/odr/TrainingStandards2005.pdf.

realities of mediator thought processes, actions, motivations, and decisions. Part I begins with a dissection of the elements of mediator neutrality. Part II introduces the science of implicit social cognition and its application to various legal contexts, turning to the mediation process as a focal point. In Part III, using one particular racial category (Asian Americans), I tease out ways in which implicit bias might affect the mediators' conclusions and actions in a particular situation.[2] Ending with Part IV, I present ideas that may help us get closer to the ideal of attaining "freedom from bias and prejudice" in mediation. I conclude that the reduction of bias and prejudice demands more attention and effort than mediators currently devote to it. We must have the intention and motivation to undertake deliberate actions to reduce unconscious bias. Bias mitigation also requires proactive steps and a more robust curriculum than what is offered in many mediation trainings, programs, and classrooms.

I. THE ESSENTIALITIES OF NEUTRALITY

Mediator neutrality is universally understood to be a vital attribute of the mediation process. The traditional definition of mediation from the 2005 revised Model Standards of Conduct for Mediations ("Model Standards"), originally approved in 1994 by the American Arbitration Association, the American Bar Association Section of Dispute Resolution, and the Association for Conflict Resolution, states, "Mediation is a process in which an impartial third party facilitates communication and negotiation and promotes voluntary decision making by the parties to the dispute."[3] Textbook definitions of the mediation process invariably use language about the involvement of a "neutral" or "impartial" third party. A sample of dispute resolution casebooks reveals similar descriptions of mediation as:

---

2. I chose Asian Americans as the focal group because of the Ds' ethnicity. Although I frame the discussion around this discrete group, I would suggest that many issues and ideas presented could be extrapolated to apply to other groups as well.

3. MODEL STANDARDS OF CONDUCT FOR MEDIATORS, Preamble (2005), *available at* http://www.abanet.org/dispute/documents/model_standards_conduct_april2007.pdf.

- "[A]n informal process in which an *impartial* third party helps others resolve a dispute or plan a transaction but does not impose a solution."[4]

- "[A] process of assisted negotiation in which a *neutral* person helps people reach agreement."[5]

- "[A] process in which a *disinterested* third party (or *'neutral'*) assists the disputants in reaching a voluntary settlement of their differences through an agreement that defines their future behavior."[6]

- "[A] process in which an *impartial* third party acts as a catalyst to help others constructively address and perhaps resolve a dispute, plan a transaction, or define the contours of a relationship."[7]

- "[A] process in which a *neutral* intervener assists two or more negotiating parties to identify matters of concern, develop a better understanding of their situation, and, based upon that improved understanding, develop mutually acceptable proposals to resolve those concerns."[8]

Neutrality is a core concept of mediation.[9] Within the profession, there is widespread consensus about the vital importance of neutrality.[10] Neutrality, along with consensuality, gives the mediation process legitimacy.[11] "The essential ingredients of classical mediation

---

4. LEONARD L. RISKIN ET AL., DISPUTE RESOLUTION AND LAWYERS 16 (4th ed. 2009) (emphasis added).
5. DWIGHT GOLANN & JAY FOLBERG, MEDIATION: THE ROLES OF ADVOCATE AND NEUTRAL 95 (2006) (emphasis added).
6. JOHN W. COOLEY, THE MEDIATOR'S HANDBOOK: ADVANCED PRACTICE GUIDE FOR CIVIL LITIGATION 2 (2000) (emphasis added).
7. CARRIE MENKEL-MEADOW ET AL., DISPUTE RESOLUTION: BEYOND THE ADVERSARIAL MODEL 266 (2005).
8. JAMES J. ALFINI ET AL., MEDIATION THEORY AND PRACTICE 1 (2d ed. 2006) (emphasis added).
9. KIMBERLEE K. KOVACH, MEDIATION: PRINCIPLES AND PRACTICE 211 (3d ed. 2004).
10. KATHERINE V.W. STONE, PRIVATE JUSTICE: THE LAW OF ALTERNATIVE DISPUTE RESOLUTION 33, 41 (2000).
11. Hilary Astor, *Rethinking Neutrality: A Theory to Inform Practice—Part I*, 11

Journal of Law & Policy   [Vol. 34:71

are: (1) its voluntariness—a party can reject the process or its outcomes without repercussions; and (2) the mediator's neutrality, or total lack of interest in the outcome."[12] As a principle "central to the theory and practice of mediation," neutrality serves "as the antidote against bias, ... [which] functions to preserve a communication context in which grievances can be voiced, claims to justice made, and agreements mutually constructed."[13]

Mediator neutrality is foundational to the mediation process. Other essential values, such as confidentiality and party self-determination, rest upon the parties' perception of the mediator as an unaligned participant. Mediator neutrality legitimizes the mediation process because the parties, rather than the mediator, are in control of decision-making.[14] To encourage the parties to share information freely and candidly with the mediator, the mediator promises not to take sides with the other party or use the information to advance the opponent's interests. Mediator neutrality makes it possible for parties to discuss issues of their choosing, negotiate with opponents, and design their own agreements.[15] Moreover, the parties' expectation of mediator neutrality is the basis upon which a relationship of trust is built.

> Trust is attained and maintained when the mediator is perceived by the disputants as an individual who understands and cares about the parties and their disputes, has the skills to guide them to a negotiated settlement, treats them impartially, is honest, will protect each party from being hurt during mediation by the other's aggressiveness or their own perceived

---

AUSTRALASIAN DISP. RESOL. J. 73, 73 (2000).

    12. COOLEY, *supra* note 6, at 2.

    13. Sara Cobb & Janet Rifkin, *Practice and Paradox: Deconstructing Neutrality in Mediation*, 16 LAW & SOC. INQUIRY 35, 35 (1991).

    14. Hilary Astor, *Rethinking Neutrality: A Theory to Inform Practice—Part II*, 11 AUSTRALASIAN DISP. RESOL. J. 145, 146 (2000).

    15. *See* Leah Wing, *Whither Neutrality?: Mediation in the Twenty-First Century*, *in* RE-CENTERING: CULTURE AND KNOWLEDGE IN CONFLICT RESOLUTION PRACTICE 93, 94 (Mary Adams Trujillo et al. eds., 2008); *see also* Scott R. Peppet, *Contractarian Economics and Mediation Ethics: The Case for Customizing Neutrality Through Contingent Fee Mediation*, 82 TEX. L. REV. 227, 256 (2003) ("[N]eutrality is considered fundamental to the self-determination for which mediation strives. To the extent that a mediator is biased towards one party, the mediator may undermine the parties' ability to craft their own solution to their problem.").

inadequacies, and has no interests that conflict with helping to bring about a resolution which is in the parties' best interest. Only when trust has been established can the parties be expected to be candid with the mediator, disclose their real interests and value the mediator's reactions . . . .[16]

Neutrality is critical to the role of the mediator.[17] Mediators must meticulously avoid even the appearance of partiality or prejudice throughout the mediation process.[18] One mediation scholar has cautioned:

Whether there is such a thing as pure neutrality or not, we know, and our clients know, that when we commit to being neutral, we are committing to not intentionally promoting one party's interests at the expense of another. When we choose to play that role, we must truly honor it, and the fact that we have a choice and decision to make about whether to put ourselves forward as a third-party neutral should only emphasize how important that commitment is.[19]

While the importance of mediator neutrality is undisputed, what actually constitutes neutrality is less clear. Neutrality is discussed, practiced, and researched rhetorically, but there are no empirical studies demonstrating exactly what neutrality means.[20] The mediator's function is nebulous due to the difficulty in defining neutrality.[21] Despite its importance, mediation literature offers slim guidance on how to achieve neutrality.[22] "Neutrality is a hard concept to nail down. It has different meanings in different cultural contexts. In some contexts, the term *neutral* is associated with being inactive,

16. NANCY ROGERS & RICHARD SALEM, A STUDENT'S GUIDE TO MEDIATION AND THE LAW 7–39 (1987), *as reprinted in* STEPHEN B. GOLDBERG ET AL., DISPUTE RESOLUTION: NEGOTIATION, MEDIATION, AND OTHER PROCESSES 113 (4th ed. 2003).
17. KOVACH, *supra* note 9, at 211.
18. COOLEY, *supra* note 6, at 28.
19. BERNARD S. MAYER, BEYOND NEUTRALITY: CONFRONTING THE CRISIS IN CONFLICT RESOLUTION 242 (2004).
20. Cobb & Rifkin, *supra* note 13, at 36–37.
21. MAYER, *supra* note 19, at 83.
22. Peppet, *supra* note 15, at 253–54.

ineffective, or even cowardly. In others, it is viewed as a sine qua non for third parties to establish respect."[23]

Comprehension of mediator neutrality is complicated by the lack of consistency in definitions. The dispute resolution lexicon is imprecise. "One reason that the theoretical concepts seem divorced from practice is that we do not yet have a shared vocabulary in our field. Although neutrality has aspects similar to fairness, justice, and appropriateness, as well as impartiality and lack of bias, it is not the same as those concepts."[24]

There is no consensus within the dispute resolution community that neutrality and impartiality are terms of art or synonyms in the vernacular.[25] Commentators and guidelines employ neutrality and impartiality circularly, asserting, for example, that "mediators shall at all times remain impartial,"[26] or "a mediator needs to remain impartial to be able to fulfill her role."[27] Neutrality and impartiality are often used synonymously when discussing a mediator's ethical duty. One reason for this is because distinctions between the terms may appear synthetic or arbitrary.[28] In their studies, Sara Cobb and Janet Rifkin found that fourteen out of fifteen mediators defined neutrality by using the word "impartiality."[29]

Other commentators and guidelines apply "neutrality" to the outcome or the elements of any resolution and "impartiality" to engagement with the parties.[30] Douglas Frenkel and James Stark propose:

---

23. MAYER, *supra* note 19, at 83.

24. Alison Taylor, *Concepts of Neutrality in Family Mediation: Contexts, Ethics, Influence, and Transformative Process*, 14 MEDIATION Q. 215, 217 (1997).

25. KOVACH, *supra* note 9, at 212 ("Neutrality is often used interchangeably with a variety of other words and phrases: *impartiality; free from prejudice or bias; not having a stake in the outcome;* and *free from conflict of interest.* Other synonyms include *unbiased, indifferent* and *independent.* There is dissention within the mediation community about whether all of these terms define neutrality, and somewhat surprisingly, whether all, or any, are appropriate characteristics for mediators.").

26. *Id.*

27. Peppet, *supra* note 15, at 264 ("I agree with the classical conception of neutrality to the extent that it recognizes the importance of impartiality.").

28. William Lucy, *The Possibility of Impartiality*, 25 OXFORD J. LEGAL STUD. 3, 13 (2005).

29. Cobb & Rifkin, *supra* note 13, at 42.

30. KOVACH, *supra* note 9, at 212–14.

"Impartiality," as we define the term, means that the mediator does not favor any one party in a mediation over any other party. Favoritism might be caused by a prior relationship or alliance with a mediation participant or by a personal bias for or against a participant based on that person's background, position, personality or bargaining style. Impartiality thus means a freedom from bias regarding the mediation *participants*.[31]

They define neutrality as meaning "that the mediator has no personal preference that the dispute be resolved in one way rather than another. The mediator is there to help the parties identify solutions that *they* find acceptable, not to direct or steer the parties toward results he favors."[32] Stated another way, neutrality is "a mediator's ability to be objective while facilitating communication among negotiating parties,"[33] and impartiality is "freedom from favoritism and bias in word, action and appearance."[34]

Despite this lack of clarity in the field, four key elements of neutrality are discernable: no conflict of interest; process equality; outcome-neutrality; and lack of bias, prejudice, or favoritism toward any party.[35] At a minimum, mediator neutrality is understood to mean

---

31. DOUGLAS N. FRENKEL & JAMES H. STARK, THE PRACTICE OF MEDIATION: A VIDEO-INTEGRATED TEXT 83–84 (2008).

32. *Id.* at 84; *see also* Susan Oberman, *Mediation Theory vs. Practice: What Are We Really Doing? Re-Solving a Professional Conundrum*, 20 OHIO ST. J. ON DISP. RESOL. 775, 802 (2000). Oberman defines impartiality as "the ability of the mediator to maintain non-preferential attitudes and behaviors towards all parties in dispute; it is the ethical responsibility of the mediator to withdraw if she or he has lost the ability to remain impartial." *Id.* She defines neutrality as the "alleged ability of the mediator to remain uninvested in the outcome of a dispute, to be aware of any contamination of neutrality, and to withdraw if he or she has lost it." *Id.*

33. Susan Nauss Exon, *The Effects that Mediator Styles Impose on Neutrality and Impartiality Requirements of Mediation*, 42 U.S.F. L. REV. 577, 580 (2008) (citing JAMES J. ALFINI ET AL., MEDIATION THEORY AND PRACTICE 12 (2001)).

34. *Id.* at 581 (quoting DISPUTE RESOLUTION ETHICS: A COMPREHENSIVE GUIDE 68 (Phyllis Bernard & Bryant Garth eds., 2002)).

35. *See* Susan Douglas, *Questions of Mediator Neutrality and Researcher Objectivity: Examining Reflexivity as a Response*, 20 AUSTRALASIAN DISP. RESOL. J. 56, 57 (2009). This study found that mediators are aware of three themes regarding neutrality and per these themes, neutrality "is understood as impartiality, even-handedness and as central to the distinction between the process and content or outcome of a dispute." *Id.* A fourth theme is also important to understanding neutrality: "'value neutrality' or the absence of a situated perspective on experience." *Id.*

that the mediator has no pecuniary interest in the subject matter, no undisclosed relationship to the parties, and no possibility of personal gain.[36] Avoiding any actual or apparent conflict of interest is subsumed in the concept of neutrality. The Uniform Mediation Act states that:

> [B]efore accepting a mediation, an individual who is requested to serve as a mediator shall: (1) make an inquiry that is reasonable under the circumstances to determine whether there are any known facts that a reasonable individual would consider likely to affect the impartiality of the mediator, including a financial or personal interest in the outcome of the mediation and an existing or past relationship with a mediation party or foreseeable participant in the mediation; and (2) disclose any such known fact to the mediation parties as soon as is practical before accepting a mediation.[37]

The Model Standards contain a similar prescription on conflicts:

> [A] mediator shall avoid a conflict of interest or the appearance of a conflict of interest during and after a mediation. A conflict of interest can arise from involvement by a mediator with the subject matter of the dispute or from any relationship between a mediator and any mediation participant, whether past or present, personal or professional, that reasonably raises a question of a mediator's impartiality.[38]

---

36. *See, e.g.,* ETHICAL GUIDELINES FOR THE PRACTICE OF MEDIATION § 4.2 (Wis. Ass'n of Mediators 1997), *available at* http://wamediators.org/pubs/ethicalquidelines.html ("As WAM members, we disclose to the parties any dealing or relationship that might reasonably raise a question about our impartiality. If the parties agree to participate in the mediation process after being informed of the circumstances, we proceed unless the conflict of interest casts serious doubt on the integrity of the process, in which case we withdraw."); *see also* COLORADO MODEL STANDARDS OF CONDUCT FOR MEDIATORS § II.A (2000), *available at* http://dola.colorado.gov/dlg/osg/docs/adrmodelstandards.pdf ("The mediator shall advise all parties of any prior or existing relationships or other circumstances giving the appearance of or creating a possible bias, prejudice, or partiality.").

37. UNIF. MEDIATION ACT §§ 9(a)(1)-(2) (2003), *available at* http://www.law.upenn.edu/bll/archives/ulc/mediat/2003finaldraft.pdf.

38. MODEL STANDARDS OF CONDUCT FOR MEDIATORS III(A) (2005), *available at* http://www.abanet.org/dispute/documents/model_standards_conduct_apn12007.

The source of the mediator's fees may compromise neutrality. A mediator must disclose any "monetary, psychological, emotional, associational, or authoritative affiliations" with any of the parties that might arguably cause a conflict of interest.[39] This aspect of neutrality has special consequences for attorney-mediators:

> One major issue for lawyers who alternate between the roles of advocate and neutral is the potential for conflicts of interest— the possibility that a party in a mediated case will be a past or future legal client of the mediator-lawyer. This is a particular concern in large law firms, where a lawyer-neutral's partners may be concerned that a single modestly compensated mediation will disqualify the entire firm from representing the party in a much more lucrative matter. Standards for neutrals call for disclosure in such situations.[40]

A second facet of neutrality is process-based or procedural, requiring that the mediator conduct the mediation process in a manner that is even-handed.[41] The Model Standards require a mediator to conduct a mediation in a manner that promotes party participation and procedural fairness.[42] "The mediator's task is to control the process of the mediation, providing a procedural framework within which the parties can decide what their dispute is about and how they wish to resolve it."[43] Process symmetry may be manifested by maneuvers such as ensuring an equal number of caucuses with the disputants or spending roughly the same amount of time with each party. It also means enforcing stated guidelines in a

---

39. KOVACH, *supra* note 9, at 213.
40. JAY FOLBERG ET AL., RESOLVING DISPUTES: THEORY, PRACTICE, AND LAW 447 (2005).
41. MODEL STANDARDS OF CONDUCT FOR MEDIATORS VI.A (2005), *available at* http://www.abanet.org/dispute/documents/model_standards_conduct_april2007.pdf ("Quality of the Process: A. A mediator shall conduct a mediation in accordance with these Standards and in a manner that promotes diligence, timeliness, safety, presence of the appropriate participants, party participation, procedural fairness, party competency and mutual respect among all participants.").
42. *Id.*
43. Hilary Astor, *Mediator Neutrality: Making Sense of Theory and Practice*, 16 SOC. & LEGAL STUD. 221, 223 (2007); *see also* Wing, *supra* note 15, at 94 ("[M]ediators are seen as only interested in the *process*, in ensuring that it is fair and that parties to the dispute are the decision-masters on any mutually acceptable agreement formulated.").

Journal of Law & Policy   [Vol. 34:71

fair manner. For example, if the mediator sets a deadline for the submission of written statements or enforces behavioral guidelines, the parties expect enforcement to be equal. "One feature of procedural impartiality is that the rules constitutive of some decision-making process must, at a minimum, favour neither party to the dispute-cum-competition or favour or inhibit both equally."[44]

Expectations of mediator neutrality encompass both procedural and outcome impartiality.[45] Neutrality in mediation is widely understood to mean that the mediator does not influence the content or outcome of the mediation. The mediator's ethical duty to be impartial throughout the process applies to her interaction with the parties and to the substance of the dispute.[46] Content-neutrality is closely linked to consensual decision-making by the disputants; it constrains mediators from usurping party control over choices and judgments.[47] Outcome neutrality requires the mediator to refrain from promoting either party's interests.[48] This component of neutrality also means the mediator should not press the parties to reach a resolution at all. "Some would draw a line at content-neutrality, however, when the result would be unfair to one of the parties or have detrimental effects on individuals with interests that are not represented at the table."[49]

A mediator's ethical duty and ability to be outcome-neutral have inspired significant debate within the profession.[50] For years, scholars

---

44. Lucy, *supra* note 28, at 11.
45. *Id.* at 8.
46. COOLEY, *supra* note 6, at 23.
47. Taylor, *supra* note 24, at 218 ("[T]he mediator is not to determine the outcome, but allow a process where decisions are made by the participants.").
48. CHRISTOPHER W. MOORE, THE MEDIATION PROCESS: PRACTICAL STRATEGIES FOR RESOLVING CONFLICT 52 (2d ed. 1996) ("What impartiality and neutrality do signify is that mediators can separate their personal opinions about the outcome of the dispute from the performance of their duties and focus on ways to help the parties make their own decisions without unduly favoring one of them.").
49. EDWARD BRUNET ET AL., ALTERNATIVE DISPUTE RESOLUTION: THE ADVOCATE'S PERSPECTIVE 200 (3d ed. 2006). In certain contexts, mediators have duties that extend beyond the immediate parties. In environmental disputes, international conflicts, and family law matters, for example, strict neutrality yields to normative consensus and standards to protect outside interests.
50. *See, e.g.*, Lawrence Susskind, *Environmental Mediation and the Accountability Problem*, 6 VT. L. REV. 1, 46–47 (1981) (asserting that environmental mediators ought to accept responsibility for ensuring that agreements are as fair and stable as possible, even though

and practitioners have questioned whether a mediator should be a mere facilitator of party-initiated outcomes or should assertively prevent agreements that are unfair or favor more powerful parties.[51] From one perspective, neutral mediators are viewed as being interested solely in ensuring a fair process, leaving the disputants to determine any mutually agreeable resolution.[52] An alternative philosophy is that mediators may or must interact with the parties unequally to account for differences such as resources, power, educational level, and financial sophistication.[53] This debate is less about how we define neutrality and more about how neutrality meshes with equally valued norms of fairness and justice, process legitimacy and quality, and party self-determination.[54] While it is important for mediators to engage in that colloquy, it is not the focus of this Article.

The final element of neutrality, and the one I want to emphasize, is the mediator's duty to "avoid bias or the appearance of bias."[55] "Impartiality between the parties and neutrality regarding the outcome are only two forms of bias. The sum total of the life experience of the mediator, the subjective self, enters into each mediation and impacts the process and outcome."[56] The Model Standards capture this in Standard II, which states in pertinent part:

"such intervention may make it difficult to retain the appearance of neutrality and the trust of the active parties"); Joseph B. Stulberg, *The Theory and Practice of Mediation: A Reply to Professor Susskind*, 6 VT. L. REV. 85, 86 (1981) ("It is precisely a mediator's commitment to neutrality which ensures responsible actions on the part of the mediator and permits mediation to be an effective, principled dispute settlement procedure."); *see also* Evan M. Rock, *Mindfulness Meditation, the Cultivation of Awareness, Mediator Neutrality and the Possibility of Justice*, 6 CARDOZO J. CONFLICT RESOL. 347, 355 (2005) (citing Peppett, *supra* note 15, at 255); Sydney E. Bernard et al., *The Neutral Mediator: Value Dilemmas in Divorce Mediation*, 4 MEDIATION Q. 61, 66 (1984).

51. *See* Carrie Menkel-Meadow, *Professional Responsibility for Third-Party Neutrals*, 11 ALTERNATIVES TO HIGH COST LITIG. 129 (1993).

52. Kimberlee K. Kovach & Lela P. Love, *Mapping Mediation: The Risks of Riskin's Grid*, 3 HARVARD NEGOT. L. REV. 71 (1998); Wing, *supra* note 15, at 94.

53. Bernard et al., *supra* note 50, at 66–67.

54. For example, family mediators must remain neutral as to outcome and impartial toward the parties but protect the best interest of children. *See* Kimberly A. Smoron, *Conflicting Roles in Child Custody Mediation: Impartiality/Neutrality and the Best Interests of the Child*, 36 FAM. & CONCILIATION CTS. REV. 258, 261 (1998).

55. Astor, *supra* note 11, at 77.

56. Oberman, *supra* note 32, at 819–20 (citing Deborah M. Kolb & Jeffrey Z. Rubin, *Mediation Through a Disciplinary Prism, in* RESEARCH ON NEGOTIATION IN ORGANIZATIONS

A. A mediator shall decline a mediation if the mediator cannot conduct it in an impartial manner. Impartiality means freedom from favoritism, bias or prejudice.

B. A mediator shall conduct a mediation in an impartial manner and avoid conduct that gives the appearance of partiality.

   1. A mediator should not act with partiality or prejudice based on any participant's personal characteristics, background, values and beliefs, or performance at a mediation, or any other reason.[57]

As of 2007, over a dozen states have implemented standards in which neutrality is defined as "freedom from favoritism or bias either by word or action, and a commitment to serve all parties as opposed to a single party."[58] Favoritism might be caused by a personal bias for or against a participant based on that person's background, position, personality or bargaining style; as such, impartiality means a freedom from bias towards the mediation participants.[59] For the disputants in mediation, a paramount concern is that the mediator has no prejudice against them on any level.[60]

To maintain neutrality, mediators must be aware of their assumptions, biases, and judgments about the participants in the process, particularly in cases where they have strong reactions to one of the parties.[61] Achieving impartiality requires mediators to have "insight into their own perspectives and experiences and [to understand] the impact that these have on their relationship with the parties in mediation."[62] "There remains the concern that the mediator's ideas and approaches to a problem will intrude and affect

---

231, 240 (Max H. Bazerman et al. eds., 3d ed. 1991)).

   57. MODEL STANDARDS OF CONDUCT FOR MEDIATORS, Standard II (2005), *available at* http://abanet.org/dispute/documents/model_standards_conduct_april2007.pdf.

   58. Exon, *supra* note 33, at 585 (quoting MINN. R. GEN. PRAC. 114 app. I cmt. 1, *available at* http://www.mncourts.gov/rules/general/GRtitleII.htm; STANDARDS OF PRACTICE: ETHICAL GUIDELINES FOR FULL MEMBERS 4 (Mont. Mediation Ass'n 1998), *available at* http://mtmediation.org/doc/Full%Ethics%20and%20Quals.pdf).

   59. FRENKEL & STARK, *supra* note 31, at 83–84.

   60. COOLEY, *supra* note 6, at 28.

   61. Taylor, *supra* note 24, at 226.

   62. Astor, *supra* note 11, at 77.

the direction of the process of mediation and its outcomes, as well as the difficulty of monitoring unconscious bias."[63]

This Article highlights the impartiality dimension of mediator neutrality in order to examine the imposing challenge presented by one form of bias,[64] i.e., implicit or unconscious bias. The next Part begins with a condensed review of the science of implicit social cognition and the phenomenon of implicit bias. It introduces the work of "behavioral realists" who import scientific research into legal analysis, and concludes with the application of these concepts to the mediation process.

II. IMPLICIT BIAS, BEHAVIORAL REALISM, AND APPLICATION TO MEDIATION

An impressive body of social science research produced over the past decades illuminates in new ways how our minds work. Advances in experimental psychology provide a deeper understanding of human perception, attention, memory, judgment, and decision-making. Cognitive social psychology studies persuasively show[65] that

---

63. *Id.*

64. There are many ways that "bias" operates in dispute resolution. *See, e.g.*, Robert S. Adler, *Flawed Thinking: Addressing Decision Biases in Negotiation*, 20 OHIO ST. J. ON DISP. RESOL. 683 (2005) (arguing that cognitive biases often associated with availability and representative and anchoring heuristics can be helpful, but can lead to stereotyping of large numbers of people based on limited past experiences; also argues that egocentric bias can affect one's perception of fairness); Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124 (1974) (contending that by understanding the positive and negative aspects of heuristics and biases, one can improve one's judgments and decisions when faced with uncertainty); John Livingood, *Addressing Bias in Conflict and Dispute Resolution Settings*, DISP. RESOL. J., Nov. 2007–Jan. 2008, at 53, 54–59 (asserting that judgment in conflict situations can be affected by four core biases: learned, incident-driven, process-driven and attributional); Joel Lee, *Overcoming Attribution Bias in Mediation: An NLP Perspective*, 15 AUSTRALASIAN DISP. RESOL. J. 48 (2004) (arguing that neuro-linguistic programming (NLP) can be useful to a mediator in helping parties understand and deal with attribution biases). A discussion of these forms of bias in mediation and negotiation is beyond the scope of this Article.

65. This research has critics and defenders. Some argue that implicit association test data do not support the conclusion that implicit bias leads to discriminatory behavior. *See* Amy L. Wax, *The Discriminating Mind: Define It, Prove It*, 40 CONN. L. REV. 979, 985 (2008) (contending that it is not "proper to equate unconsciously biased mental associations with the tendency to engage in unlawful discrimination"); R. Richard Banks, Jennifer L. Eberhardt & Lee Ross, *Discrimination and Implicit Bias in a Racially Unequal Society*, 94 CALIF. L. REV. 1169, 1187–88 (2006) (asserting that the Implicit Association Test (IAT) is not significantly

unconsciously held attitudes and stereotypes can affect our interaction with others and may predict behavior.[66] This rich reservoir of scientific material deserves a more expansive presentation than I am able to offer here. What follows is a selective summary of some of the fascinating, and often startling, experimental discoveries about the insidious operation of unconscious bias. In the interest of space, I omit detailed descriptions of experimental design and administration and refer readers to the sources for explanations of methodologies and statistical analyses.

Following this summary of implicit bias research, I present the work of "behavioral realists." These legal academics and social scientists use social cognition research to measure how legal doctrines and institutional processes address discriminatory behavior. In contexts such as peremptory challenges, judicial decision-making, employment, and jury selection, scholars argue that current procedural and substantive legal protections fail to account for the

---

correlated to discriminatory behavior because subtle behaviors such as eye contact, speech errors, and body language do not constitute discriminatory action); Philip E. Tetlock, *Cognitive Biases and Organizational Correctives: Do Both Disease and Cure Depend on the Politics of the Beholder?*, 45 ADMIN. SCI. Q. 293 (2000) (arguing that studies should not focus on judgmental shortcomings but on the fact that everyone cannot fit in a particular category, and that an ideological bias on the part of researchers does not always translate to a "real-world" setting); Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023 (2006) (claiming that implicit bias research is invalid and should not be used in developing antidiscrimination law). There are rebuttals to this criticism. *See* Samuel R. Bagenstos, *Implicit Bias, "Science," and Antidiscrimination Law*, 1 HARV. L. & POL'Y REV. 477 (2007) (discrediting critics such as Mitchell and Tetlock for dismissing research unscientifically and subjectively, and further arguing that sufficient evidence exists to show that implicit biases lead to discrimination, and that antidiscrimination laws should be used to counter implicit bias effects); David L. Faigman et al., *A Matter of Fit: The Law of Discrimination and the Science of Implicit Bias*, 59 HASTINGS L.J. 1389, 1389–99, 1426–29 (2007) (arguing that expert testimony regarding research on implicit bias should be admissible in Title VII discrimination cases as a general background of implicit bias to give triers of fact understanding and context because "studies using a variety of measures and techniques have demonstrated the effects of implicit bias on judgments and behavior, creating a broad research base that spans several social scientific disciplines including psychology, sociology, and organizational behavior"; therefore "it is a mistake to conflate the existence of implicit bias with any one measure such as the IAT," or Implicit Association Test, and "it is a mistake to assume that critiques of one particular measure such as the IAT undermine the entire body of evidence showing the existence of implicit stereotypes and bias and their impact on judgments and behavior in the workplace").

    66. Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945, 954–55 (2006).

operation of unconscious biases. With evidence that implicit attitude measures reveal much more bias favoring advantaged groups than do explicit measures, adherents of behavioral realism advocate legal reform to adequately address prejudiced behavior. I examine the mediation process through a behavioral realism lens and suggest that mediators regularly fail to act in unbiased ways.

A. Implicit Bias Research

Implicit social cognition is "a broad theoretical category that integrates and reinterprets established research findings, guides searches for new empirical phenomena, prompts attention to presently undeveloped research methods, and suggests applications in various practical settings."[67] Implicit social cognitionists posit that we can learn more about stereotypes and prejudice when we examine their unconscious operations. For example, experiments examining the causal relationship between unconscious stereotypes and biases in perception and memory have shined new light on social interactions and led theorists to recommend corrective actions to counteract the pervasiveness of unconscious biases.[68] Mental processes such as implicit memory, implicit attitudes, implicit self-esteem, implicit perception, and implicit stereotypes operate outside conscious attention and thereby unconsciously influence judgment.[69] "The term *implicit*, contrasted with *explicit*, is used to capture a distinction variously labeled as unconscious versus conscious, unaware versus aware, and indirect versus direct."[70] The most commonly used techniques for studying implicit social cognition are priming tasks with rapid response time measures and the Implicit Association Test (IAT), which is described below.[71]

---

67. Anthony G. Greenwald & Mahzarin R. Banaji, *Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes*, 102 PSYCHOL. REV. 4, 4 (1995).
68. Mahzarin R. Banaji & Anthony G. Greenwald, *Implicit Stereotyping and Prejudice, in* 7 THE PSYCHOLOGY OF PREJUDICE: THE ONTARIO SYMPOSIUM 55, 56 (Mark P. Zanna & James M. Olson eds., 1994).
69. Greenwald & Krieger, *supra* note 66, at 947.
70. Mahzarin R. Banaji, Curtis Hardin & Alexander J. Rothman, *Implicit Stereotyping in Person Judgment*, 65 J. PERSONALITY & SOC. PSYCHOL. 272 n.1 (1993).
71. Kristin A. Lane, Jerry Kang & Mahzarin R. Banaji, *Implicit Social Cognition and Law*, 3 ANN. REV. L. & SOC. SCI. 427, 428, 431 (2007).

Implicit bias refers to:

> [A]n aspect of the new science of unconscious mental processes that has substantial bearing on discrimination law. Theories of implicit bias contrast with the "naïve" psychological conception of social behavior, which views human actors as being guided solely by explicit beliefs and their conscious intentions to act. A belief is *explicit* if it is consciously endorsed. An intention to act is *conscious* if the actor is aware of taking an action for a particular reason. . . . In contrast, the science of implicit cognition suggests that actors do not always have conscious, intentional control over the processes of social perception, impression formation, and judgment that motivate their actions.[72]

An overview of implicit social cognition research draws four main conclusions about the collective findings: (1) there is a variance, sometimes wide, between implicit and explicit cognition; (2) there is a discernable, pervasive and strong favoritism for one's own group, as well as for socially valued groups; (3) implicit cognitions, often more accurately than explicit, predict behavior; (4) implicit social cognitions are not impervious to change.[73]

Two concepts are key to the study of implicit social cognition: attitude (or preference) and stereotype (or belief).[74] Attitudes can be defined as dispositions toward things, such as people, places, and policies.[75] Stated another way, "an *attitude* [is] an evaluative disposition—that is, the tendency to like or dislike, or to act favorably or unfavorably toward, someone or something."[76] Explicit attitude expression can come in the form of action, such as selecting something we like or rejecting something we dislike.[77] Implicit attitudes are "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable

---

72. Greenwald & Krieger, *supra* note 66, at 946.
73. Lane, Kang & Banaji, *supra* note 71, at 431–38.
74. *Id.* at 429.
75. Greenwald & Banaji, *supra* note 67, at 7.
76. Greenwald & Krieger, *supra* note 66, at 948.
77. *Id.*

feeling, thought, or action toward social objects."[78] For example, "[a]n implicit attitude toward B may be indirectly indicated by a (direct) measure of evaluation of A, when A and B have some relation that predisposes the implicit influence."[79] "Halo effect" research provides another example: physically attractive men and women "are judged to be kinder, more interesting, more sociable, happier, stronger, of better character, and more likely to hold prestigious jobs" by operation of an "objectively irrelevant attribute [physical attractiveness] that influences evaluative judgment on various other dimensions."[80]

A stereotype "is a mental association between a social group or category and a trait."[81] Stereotyping is "the application of beliefs about the attributes of a group to judge an individual member of that group."[82] A person's attitude toward someone or something is a consistent positive or negative response to an object.[83] On the other hand,

> a stereotype may encompass beliefs with widely diverging evaluative implications. For example, the stereotype of members of a certain group (e.g., cheerleaders) may

---

78. Greenwald & Banaji, *supra* note 67, at 8.
79. *Id.*
80. *Id.* at 9 (citing Karen Dion, Ellen Berscheid & Elaine Walster, *What is Beautiful is Good*, 24 J. PERSONALITY & SOC. PSYCHOL. 207 (1972)). The act of voting presents another example of implicit attitude. Voting for Obama because you know you like his beliefs and policies would be an explicit attitude expression. However, "a vote might function as an *implicit attitude indicator*—that is, an action that indicates favor or disfavor toward some object but is not understood by the actor as expressing that attitude. For example, a voter may vote for a particular candidate even though the voter knows nothing other than the candidate's name shares initial letters with the voter's name. In such a case, the vote can be understood, at least in part, as an implicit expression of the voter's self-favorable attitude." Greenwald & Krieger, *supra* note 66, at 948. Reliable research finds that most people have a positive attitude about themselves. Thus, "an expectable form of implicit attitude effect is that novel objects that are invested with an association to self should be positively evaluated." Greenwald & Banaji, *supra* note 67, at 10. Continuing with the voting example, even if you know nothing about Obama's sister, you might like his sibling. "This favorable attitude is an *implicit indicator* of attitude toward the candidate. Here, the 'implicit' designation indicates that the attitude expressed toward the candidate determined the attitude toward the relative, even though the liking or disliking for the relative may be experienced as an independent attitude." Greenwald & Krieger, *supra* note 66, at 948–49.
81. Greenwald & Krieger, *supra* note 66, at 949.
82. Banaji & Greenwald, *supra* note 68, at 58.
83. Greenwald & Banaji, *supra* note 67, at 7.

simultaneously include the traits of being physically attractive
(positive) and unintelligent (negative). Stereotypes guide
judgment and action to the extent that a person acts toward
another as if the other possesses traits included in the
stereotype.[84]

Stereotypes are activated automatically, generally leading to the
presumption that "the operation of the stereotype or prejudice [is]
unintended by the research participants (i.e., not deliberate), either
because they are unaware of certain critical aspects of the procedure
or because they are operating under conditions that make it difficult
to deliberately base responses on specific beliefs or evaluations."[85]
For example, a 1983 experiment conducted by Samuel Gaertner and
John McLaughlin provided one illustration of stereotype activation,
demonstrating that subjects more quickly identified word pairs if they
were consistent rather than inconsistent with African American
stereotypes (e.g., Blacks-lazy vs. Blacks-ambitious).[86]

More recently, Mahzarin Banaji and Curtis Hardin conducted two
priming task experiments on gender stereotyping.[87] Subjects saw
gender-related primes (e.g., mother, father) or neutral primes (e.g.,
parent, student) followed by target words. Subjects in the first
experiment were asked to respond as to whether the following target
pronoun, either gender-related (e.g., he, she) or neutral (e.g., it, me),
was male or female. Participants were able to respond faster to
pronouns that were consistent with the gender stereotype of the
prime; this result occurred independently of explicit beliefs about
gender stereotypes.[88] The second experiment asked participants only
to identify whether the target word was a pronoun or not a pronoun,
but still resulted in similar effects of gender stereotyping.[89] These

---

84. *Id.* at 14.
85. Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6
PERSONALITY & SOC. PSYCHOL. REV. 242, 243 (2002).
86. *Id.* at 242 (citing Samuel L. Gaertner & John P. McLaughlin, *Racial Stereotypes:
Associations and Ascriptions of Positive and Negative Characteristics*, 46 SOC. PSYCHOL. Q. 23
(1983)).
87. *See* Mahzarin R. Banaji & Curtis D. Hardin, *Automatic Stereotyping*, 7 PSYCHOL. SCI.
136 (1996).
88. *Id.* at 136–39.
89. *Id.* at 139–40.

experiments "demonstrated that judgments of targets that follow[ed] gender-congruent primes are made faster than judgments of targets that follow[ed] gender-incongruent primes," showing that gender information imparted by words can automatically influence judgment, even in unrelated tasks.[90] Other studies bolster the finding that "[p]eople may often not be aware of what they are doing, they might even intend to be doing something else; perhaps worst of all, the operation of stereotypes and prejudice may be outside of their control."[91]

Automatic activation of stereotypes "provides the basis for implicit stereotyping."[92] "*Implicit stereotypes* are the introspectively unidentified (or inaccurately identified) traces of past experience that mediate attributions of qualities to members of a social category."[93] In one study, Mahzarin Banaji and Anthony Greenwald examined the relationship between implicit stereotypes and gender.[94] When testing participants' recognition of famous names, participants were more likely to falsely identify a male name as famous than they were to falsely identify a female name as famous. The false-fame effect was substantial when the names were male but weaker when the names were female, demonstrating an implicit indicator of the stereotype that associates maleness with fame (and achievement).[95] Researchers observe that stereotypes are often expressed implicitly in the behavior of people who expressly disavow the stereotype. Because race and gender stereotypes have been studied more often, they provide the "most persuasive evidence for implicit stereotyping."[96]

"*Implicit biases* are discriminatory biases based on implicit attitudes or implicit stereotypes. Implicit biases are especially

---

90. *Id.* at 140. In another experiment, researchers discovered that by activating abstract knowledge about beliefs associated with men and women, such as dependence and aggressiveness, subjects judged male and female targets more harshly when the targets' group membership stereotypically matched (e.g., after the subject's exposure to dependence primes, the subject will judge the female target to be more dependent). Banaji, Hardin & Rothman, *supra* note 70, at 272.

91. Blair, *supra* note 85, at 242.

92. Greenwald & Banaji, *supra* note 67, at 15.

93. *Id.*

94. Mahzarin R. Banaji & Anthony G. Greenwald, *Implicit Gender Stereotyping in Judgments of Fame*, 68 J. PERSONALITY & SOC. PSYCHOL. 181 (1995).

95. *Id.*

96. Greenwald & Banaji, *supra* note 67, at 15.

Journal of Law & Policy   [Vol. 34:71

intriguing, and also especially problematic, because they can produce behavior that diverges from a person's avowed or endorsed beliefs or principles."[97] The existence of stereotypes and biases does not mean that a person necessarily holds consciously prejudicial beliefs. Stereotypes and prejudices unconsciously and naturally form "through ordinary biases rooted in memory" to simplify cognitive processes.[98] To a varying degree, all of us are subject to the operation of implicit stereotyping and prejudice.[99] "The best of intentions do not and cannot override the unfolding of unconscious processes, for the triggers of automatic thought, feeling, and behavior live and breathe outside conscious awareness and control."[100]

In large part, implicit social cognition research has advanced because of the development and accessibility of the Implicit Association Test (IAT), an instrument that produces an implicit-attitude measure based on response speeds in two four-category tasks.[101] Since 1998, self-administered IAT demonstrations have been available online.[102] The most widely used version is the "Race IAT" which measures implicit attitudes toward African Americans (AA) relative to European Americans (EA).[103]

Using the IAT, social scientists have found that most Americans exhibit a "strong and automatic positive evaluation of White

---

97. Greenwald & Krieger, *supra* note 66, at 951.
98. Mahzarin R. Banaji & R. Bhaskar, *Implicit Stereotypes and Memory: The Bounded Rationality of Social Beliefs, in* MEMORY, BRAIN, AND BELIEF 139, 167 (Daniel L. Schacter & Elaine Scarry eds., 2000).
99. *Id.* at 143.
100. *Id.* at 142–43.
101. *See* Anthony G. Greenwald, Mahzarin R. Banaji & Brian A. Nosek, *Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm*, 85 J. PERSONALITY & SOC. PSYCHOL. 197 (2003).
102. PROJECT IMPLICIT, https://implicit.harvard.edu/implicit/ (last visited Nov. 6, 2010).
103. The IAT works as follows: "[R]espondents first practice distinguishing AA from EA faces by responding to faces from one of these two categories with the press of a computer key on the left side of the keyboard and to those of the other category on the right side of the keyboard. Respondents next practice distinguishing pleasant-meaning from unpleasant-meaning words in a similar manner. The next two tasks, given in a randomly determined order, use all four categories (AA faces, EF faces, pleasant-meaning words, and unpleasant-meaning words). In one of these two tasks, the IAT calls for one response (say, pressing a left-side key) when the respondent sees AA faces or pleasant words, whereas EA faces and unpleasant words call for the other response (right-side key). In the remaining task, EA faces share a response with pleasant words and AA faces with unpleasant words." Greenwald & Krieger, *supra* note 66, at 952–53.

Americans and a relatively negative evaluation of African Americans."[104]

An analysis of data archived from many years of web-accessed IAT interactive demonstrations compared the level of favoritism toward advantaged versus disadvantaged groups revealed by implicit and explicit measures. Over two million people have taken the IAT; 90 percent have been American.[105] Eighty-eight percent of white test takers have manifested implicit bias in favor of Whites and against Blacks.[106] Over 80 percent of heterosexuals manifested implicit bias in favor of straights over gays and lesbians.[107] Non-Arab and non-Muslim test takers manifested strong implicit bias against Muslims.[108] These results are in sharp contrast to self-reported attitudes.[109] The following generalizations are apparent as to these self-selected users: explicit measures show much greater evidence for attitudinal impartiality or neutrality, and the IAT measures revealed greater bias in favor of the advantaged group. Implicit attitude measures reveal far more bias favoring advantaged groups than do explicit measures.[110] Interestingly, only African Americans failed to show substantial pro-EA race bias on the Race IAT.[111] From this, one can draw the conclusion that "*any* non-African American subgroup of the United States population will reveal high proportions of persons showing statistically noticeable implicit race bias in favor of EA relative to AA."[112]

Becca Levy and Mahzarin Banaji surveyed research that utilized the IAT and implicit priming to measure automatic attitudes and stereotypes related to age.[113] Based on 68,144 tests that included people along a wide spectrum of ages, Levy and Banaji offered three

---

104.   Nilanjana Dasgupta et al., *Automatic Preference for White Americans: Eliminating the Familiarity Explanation*, 36 J. EXPERIMENTAL SOC. PSYCHOL. 316, 316 (2000).

105.   Shankar Vedantam, *See No Bias*, WASH. POST MAG., Jan. 23, 2005, at 12, 15.

106.   *Id.*

107.   *Id.*

108.   *Id.*

109.   *Id.*

110.   Greenwald & Krieger, *supra* note 66, at 955.

111.   *Id.* at 956.

112.   *Id.*

113.   Becca R. Levy & Mahzarin R. Banaji, *Implicit Ageism, in* AGEISM: STEREOTYPING AND PREJUDICE AGAINST OLDER PERSONS 49, 51–52 (Todd D. Nelson ed., 2002).

key findings.[114] First, ageism, defined as "an alteration in feeling, belief, or behavior in response to an individual's or group's perceived chronological age[,] . . . can operate without conscious awareness, control, or intention to harm."[115] Levy and Banaji found implicit ageism to be among the largest negative implicit attitudes observed, even larger than the anti-black attitude among white Americans.[116] Second, explicit age attitudes toward the elderly are negative, but implicit age attitudes are far more negative overall.[117] Third, a peculiar feature of implicit ageism is that it does not appear to vary as a function of age, since both older and younger subjects tend to have negative implicit attitudes toward the old and positive implicit attitudes toward the young.[118] The authors argue that ageism occurs implicitly and that all people are implicated in it. "Once age stereotypes have been acquired, they are likely to be automatically triggered by the presence of an elderly person."[119]

When implicit and explicit attitudes toward the same object vary, the discrepancy between the two is referred to as dissociation. This is often seen in attitudes toward stigmatized groups defined by age, race, sexual orientation, and disability.[120] Experiments show that implicit expressions of beliefs and attitudes are unrelated to explicit versions of the same. Two studies explored the use of the IAT "to chart the emergence of implicit attitudes in early and middle childhood."[121] The first study examined white American children's attitudes of blacks and Japanese.[122] The second also tested for explicit and implicit race biases but used a sample from a rural Japanese town where participants had little exposure to out-groups.[123] Generally, implicit and explicit biases existed at the earliest ages tested, but dissociation began around age ten or middle childhood as

---

114. *Id.* at 54.
115. *Id.* at 50.
116. *Id.* at 54–55.
117. *Id.* at 55.
118. *Id.*
119. *Id.* at 64.
120. Greenwald & Krieger, *supra* note 66, at 949.
121. Yarrow Dunham et al., *From American City to Japanese Village: A Cross-Cultural Investigation of Implicit Race Attitudes*, 77 CHILD DEV. 1268, 1270 (2006).
122. *Id.* at 1270–71.
123. *Id.* at 1274.

participants' explicit bias began to dissipate.[124] Researchers consistently observed dissociation between conscious and unconscious social judgment.[125]

Significantly, implicit bias predicts individually discriminatory behaviors.[126] Studies substantiate that "implicit measures of bias have relatively greater predictive validity than explicit measures in situations that are socially sensitive, like racial interactions, where impression-management processes might inhibit people from expressing negative attitudes or unattractive stereotypes."[127] An experiment featuring doctors making patient assessments provides an example of discriminatory behavior predicted by implicit bias measures.[128] Physicians with stronger implicit anti-black attitudes and stereotypes were not as likely to prescribe a medical procedure for African Americans compared to white Americans with the same medical profiles.[129] In addition, implicit measures are relatively better predictors of "spontaneous behaviors such as eye contact, seating distance, and other such actions that communicate social warmth or discomfort."[130] "Those who possess stronger negative attitudes toward a stigmatized group tend to exhibit more negative behaviors (e.g., blinking) and less positive behaviors (e.g., smiling) when interacting with a member of that group."[131]

Researchers conclude:

> The exposure of stereotyped knowledge in these studies represents an experimental analog of the countless ways in everyday life by which stereotyped information is continuously made available. . . . [I]mplicit stereotyping effects undermine the current belief about the role of consciousness in guaranteeing equality in the treatment of individuals irrespective of sex, class, color, and national origin. . . . Implicit stereotyping critically compromises the efficacy of

---

124. *Id.* at 1270, 1274–76.
125. Banaji & Bhaskar, *supra* note 98, at 146.
126. Lane, Kang & Banaji, *supra* note 71, at 436.
127. Greenwald & Krieger, *supra* note 66, at 954–55.
128. Lane, Kang & Banaji, *supra* note 71, at 430.
129. *Id.*
130. Greenwald & Krieger, *supra* note 66, at 955.
131. Lane, Kang & Banaji, *supra* note 71, at 436.

"good intention" in avoiding stereotyping and points to the importance of efforts to change the material conditions within which (psychological) stereotyping processes emerge and thrive.[132]

B. Behavioral Realism

With so much laboratory evidence to support findings in implicit social cognition, many commentators have argued that we should consider the legal implications of this new science.[133] Over twenty years ago legal scholar Charles Lawrence called attention to the effects of unconscious racism in an oft-cited law review article, noting that "a large part of the behavior that produces racial discrimination is influenced by unconscious racial motivation."[134] Social science research has spawned a new generation of academics who question whether existing legal doctrines realistically account for the operation of implicit social cognition on human actors.[135]

---

132. Banaji, Hardin & Rothman, *supra* note 70, at 280.

133. Several authors have surveyed research and experiments on metacognitive processes to show how awareness, control, and intentionality (features of consciousness) relate to the formation of beliefs, attitudes, and behaviors. They argue that research on implicit social processes, particularly data on influences outside conscious awareness, control, and intention, may drive re-conceptualization of the legal notion of intention as it relates to discrimination. *See, e.g.*, Lane, Kang & Banaji, *supra* note 71; Banaji & Bhaskar, *supra* note 98; Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CALIF. L. REV. 997 (2006); Mahzarin R. Banaji & Nilanjana Dasgupta, *The Consciousness of Social Beliefs: A Program of Research on Stereotyping and Prejudice*, *in* METACOGNITION: COGNITIVE AND SOCIAL DIMENSIONS 157, 167 (Vincent Y. Yzerbyt et al. eds., 1998).

134. Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317, 322 (1987).

135. *See generally* Jennifer S. Hunt, *Implicit Bias and Hate Crimes: A Psychological Framework and Critical Race Theory Analysis*, *in* SOCIAL CONSCIOUSNESS IN LEGAL DECISION MAKING: PSYCHOLOGICAL PERSPECTIVES 247, 255 (Richard L. Wiener et al. eds., 2007) (arguing that implicit stereotypes and prejudice may "tip the scale" in triggering hate crimes by causing hostile interpretations, increasing the likelihood of categorizing an individual as a member of a stigmatized group, activating aggressive behavioral tendencies, and/or lowering the decision threshold for aggressive behavior); Antony Page, *Unconscious Bias and the Limits of Director Independence*, 2009 U. ILL. L. REV. 237 (arguing that rules regarding director independence are flawed because they do not account for sources of bias, especially unconscious bias); Sara R. Benson, *Reviving the Disparate Impact Doctrine to Combat Unconscious Discrimination: A Study of* Chin v. Runnels, 31 T. MARSHALL L. REV. 43, 58–59 (2005) (arguing that the intent doctrine should be struck and the disparate impact doctrine should be reinstated in Equal Protection cases to combat implicit discrimination).

In *Trojan Horses of Race*, an exposition on selected findings in social cognition research, Jerry Kang describes "'racial mechanics'— the ways in which race alters intrapersonal, interpersonal, and intergroup interactions."[136] With an emphasis on implicit bias material, Kang urges that "it is time for a new 'behavioral realist' approach, which draws on the traditions of legal realism and behavioral science."[137] The term "behavioral realism" was coined by a collection of academics to identify a collaboration of legal scholars and social cognitionists that "seeks to apply the best model of human behavior that science has made available to questions of law and policy."[138] The idea of behavioral realism is that law and jurisprudence should be consistent with accepted interpretations of behavioral science.[139] One example of this type of collaboration is Kang and Banaji's proposal to apply implicit social cognition research to create a new framework for affirmative action, using a methodology that "forces the law to confront an increasingly accurate description of human decision making and behavior, as provided by the social, biological, and physical sciences."[140] Kang and Banaji contend, "[b]ehavioral realism identifies naïve theories of human behavior . . . [and] juxtaposes these theories against the best scientific knowledge available to expose gaps between assumptions embedded in law and reality described by science. When behavioral realism identifies a substantial gap, the law should be changed to comport with science."[141]

A number of scholars have employed a behavioral realist approach to evaluate legal doctrines that require a showing of explicit

---

136. Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1493 (2005).
137. *Id.* at 1494 n.21.
138. *Id.*
139. *See, e.g.*, Dale Larson, *Unconsciously Regarded as Disabled: Implicit Bias and the Regarded-As Prong of the Americans with Disabilities Act*, 56 UCLA L. REV. 451, 476, 484–87 (2008) (citing a study that found "[p]reference for people without disabilities compared to people with disabilities was among the strongest implicit and explicit effects across the social group domains," and concluding that amendments to the Americans with Disabilities Act, which would reinstate a broader definition of a key element of actionable discrimination, are an important step forward in protecting against disability discrimination resulting from implicit bias).
140. Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of "Affirmative Action,"* 94 CALIF. L. REV. 1063, 1064–65 (2006).
141. *Id.* at 1065.

Journal of Law & Policy   [Vol. 34:71

bias and conscious racial motivation. In the area of employment discrimination law, Linda Krieger and Susan Fiske assert that requirements based on intentionality and consciously discriminatory motivations are out of sync with empirical data from psychological science.[142] Relying on studies showing commonly held gender stereotypes and research indicating that implicit stereotypes remain in people who expressly hold egalitarian views, David Faigman, Nilanjana Dasgupta, and Cecilia Ridgeway argue that employment discrimination law requires new interpretations relying on more than explicit motivations.[143]

In articles addressing juror and judicial decision-making, authors present scientific research to show that implicit bias affects courtroom proceedings, suggesting that judges who prohibit references to race or other social characteristics during the proceedings are actually allowing discrimination to continue rather than helping to stop it.[144] Judges who strive to create a prejudice-free courtroom face an additional quandary. Studies confirm that unconscious bias may explain, at least in part, disparities in judicial decision-making, such as with convictions and sentencing.[145] Concerned with the impact of implicit bias in the process of creating a fair cross-section of jurors, one judge recognized that racial dynamics played out in jury deliberations, but she was frustrated in her attempts to remove prejudiced jurors from the pool.[146] Looking at

---

142. Krieger & Fiske, *supra* note 133, at 1061–62.

143. Faigman et al., *supra* note 65, at 1434 (concluding that expert testimony regarding research on implicit bias should be admissible in Title VII discrimination cases to provide a general background of implicit bias and give triers of fact understanding and context, but not for testimony that implicit bias influenced an employment decision in a specific case).

144. *See* Jody Armour, *Stereotypes and Prejudice: Helping Legal Decisionmakers Break the Prejudice Habit, in* CRITICAL RACE REALISM: INTERSECTIONS OF PSYCHOLOGY, RACE, AND LAW 11 (Gregory S. Parks et al. eds., 2008).

145. *See* Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195, 1202 (2009). The authors found that the white judges in their study may have been compensating for unconscious racial biases in their decision-making, at least when the defendant's race was clearly identified. *Id.* at 1223. However, the black judges in the study had a greater propensity to convict the African American defendant, perhaps, as the authors speculate, because "[b]lack judges . . . might have been less concerned with appearing to favor the black defendant than the white judges." *Id.* at 1224.

146. Janet Bond Arterton, *Unconscious Bias and the Impartial Jury*, 40 CONN. L. REV. 1023, 1030 (2008) ("The harsh reality for judges conducting voir dire aimed at seating only fair and impartial jurors is that the jurors themselves may not be able to assist."); *see also* Turner v.

peremptory challenges, Anthony Page argues that the current three-step *Batson* approach[147] is inadequate to address the phenomenon of racially motivated challenges in jury selection.[148] The *Batson* approach requires that the challenging lawyer actually be conscious of her reason for striking, but research shows that unconscious bias can easily alter our perceptions of others.[149] Page's piece, along with other social science articles, was cited by Justice Breyer in *Miller-El v. Dretke*, a case in which the Supreme Court concluded that a prosecutor's use of peremptory challenges to strike several black jurors constituted purposeful discrimination.[150] Justice Breyer commented that "[s]ubtle forms of bias are automatic, unconscious, and unintentional,"[151] operating outside the knowledge of the person acting in a biased manner.

C. Application to Mediation

Unlike judges, mediators lack the authority to render binding judgments. Nevertheless, they may have significant influence on individual lives. A mediator's actions, judgments, strategic choices, and interactions with the disputants have an undeniable impact on the substance of the mediation and the results of the mediation process. In her book on mediator behavior, Deborah Kolb described her

---

Stime, 222 P.3d 1243 (Wash. 2009) (holding that the jurors' racially biased conduct in regards to a Japanese lawyer supported grounds for a new trial); Martha Neil, *New Trial Sought After Jurors Mock Lawyer's Heritage*, ABA JOURNAL (Jan. 15, 2008, 4:34 PM), http://www.aba journal.com/news/new_trial_sought_after_jurors_mock_lawyers_heritage (Washington lawyer sought new trial after jurors mocked his Japanese heritage during deliberations).

147. In *Batson v. Kentucky*, 476 U.S. 79 (1986), the Supreme Court provided a three step approach for constitutional claims regarding the use of peremptory challenges. The first step requires the defendant to raise the inference that the prosecutor used peremptory challenges to exclude possible jurors based on race. *Id.* at 96. In the second step, the prosecution has the burden of producing a race-neutral explanation for the exclusion of the jurors. *Id.* at 97. In the third step, the trial court must determine if the defendant has proven purposeful discrimination. *Id.* at 98.

148. Antony Page, *Batson's Blind-Spot: Unconscious Stereotyping and the Peremptory Challenge*, 85 B.U. L. REV. 155 (2005).

149. "[T]he problem with *Batson* is its inability to address the honest, well-intentioned lawyer who nevertheless still discriminates." *Id.* at 179 (emphasis added). The lawyer's lack of self-awareness may lead to peremptory challenges being exercised in a discriminatory manner even though the lawyer states, and believes, she has a non-discriminatory reason. *Id.* at 234–35.

150. Miller-El v. Dretke, 545 U.S. 231, 265–66 (2005).

151. *Id.* at 268 (Breyer, J., concurring) (internal quotations omitted).

Journal of Law & Policy   [Vol. 34:71

observations of labor mediators during several mediations.[152] She observed two contrasting types of mediator behavior, leading her to classify mediators as either "orchestrators" or "dealmakers".[153] Orchestrators tended to require that the parties take more responsibility for negotiating, designing settlement proposals, and convincing their colleagues to accept a given settlement.[154] Dealmakers, on the other hand, saw themselves as responsible for creating, pushing, and "selling" an ultimate settlement to the parties.[155] Mediators in Kolb's study admitted to "manipulat[ing]" the parties to certain outcomes.[156] Kolb observed mediators using "direct persuasion . . . resulting in a deal that bears the imprint of the mediator as much as it does the parties."[157]

This spectrum of mediator behavior has been described in various ways. Leonard Riskin's well-known grid situates mediators within a "facilitative-evaluative/broad-narrow" framework.[158] Ellen Waldman uses "Norm-Generating," "Norm-Educating," and "Norm-Advocating" terminology.[159] Hilary Astor compares a "robust" approach, in which the mediator is "assertive, active, and interventionist," to a "minimalist" approach that entails convening, stimulating information flow, and identifying options.[160] For every mediator who argues that a facilitative model is the better or "correct" approach, another advocates a more directive approach in fulfilling duties.[161] By analyzing mediators in practice, observers

---

152. DEBORAH M. KOLB, THE MEDIATORS (1983).
153. *Id.* at 25.
154. *Id.* at 34–41, 42–43.
155. *Id.* at 34–42.
156. *Id.* at 41.
157. *Id.* at 42.
158. Leonard L. Riskin, *Understanding Mediators' Orientations, Strategies, and Techniques: A Grid for the Perplexed*, 1 HARV. NEGOT. L. REV. 7, 16–35 (1996); Leonard L. Riskin, *Decisionmaking in Mediation: The New Old Grid and The New New Grid System*, 79 NOTRE DAME L. REV. 1, 12–13 (2003) (proposing substituting "directive" and "elicitive" for "evaluative" and "facilitative").
159. Ellen A. Waldman, *The Challenge of Certification: How to Ensure Mediator Competence While Preserving Diversity*, 30 U.S.F. L. REV. 723, 728–43 (1996).
160. Astor, *supra* note 11, at 75–76.
161. *Compare* Kimberlee K. Kovach & Lela P. Love, *"Evaluative" Mediation is an Oxymoron*, 14 ALTERNATIVES TO HIGH COST LITIG. 31 (1996) ("An essential characteristic of mediation is facilitated negotiation. . . . 'Evaluative' mediation is an oxymoron. It jeopardizes neutrality because a mediator's assessment invariably favors one side over the other."), *with*

have concluded that evaluative mediators cross the neutrality line in ways that facilitative practitioners do not.[162] It is when mediators move from "educative" and "rational-analytic" roles to "therapeutic" and "normative-evaluative" roles "that an ethics dilemma regarding neutrality and impartiality may arise."[163]

Exoneration of facilitative mediators from neutrality breaches, however, may be too generous. Under the assumption that "mediators themselves routinely and unabashedly engage in manipulation and deception to foster settlements," James Coben argues that "[t]his is not simply a matter of mediator style—the [much-discussed] distinction between facilitative and evaluative approaches."[164] Despite neutrality constraints, Coben asserts that mediators "*are directly involved in influencing disputants toward settlement*."[165]

Mediator partiality is manifested in subtle ways.[166] Two studies reveal a significant disconnect between the articulated practice goal of neutrality and the actual techniques and strategies of mediators. In the first study, empirical research into community mediation in neighbor disputes showed that mediators (paid staff and trained volunteers) found it difficult to ignore "personal bias and evaluations of the worthiness of particular claims and disputants."[167] Mediators confessed to being so angry or frustrated with a disputant that on occasion "they felt they could not even make a pretence at remaining neutral."[168] Instead of being a rare occurrence, mediators stated their

Donald T. Weckstein, *In Praise of Party Empowerment—and of Mediator Activism*, 33 WILLAMETTE L. REV. 501, 504 (1997) ("When consistent with the parties' expectations and the mediator's qualifications, activist intervention by the mediator should be encouraged rather than condemned.").

162. Linda Mulcahy, *The Possibilities and Desirability of Mediator Neutrality—Towards an Ethic of Partiality?*, 10 SOC. & LEGAL STUD. 505, 510–11 (2001).

163. Taylor, *supra* note 24, at 221.

164. James R. Coben, *Mediation's Dirty Little Secret: Straight Talk About Mediator Manipulation and Deception*, 2 ALTERNATIVE DISP. RESOL. EMP. 4 (2004).

165. *Id.* at 5 (citing CHRISTOPHER MOORE, THE MEDIATION PROCESS: PRACTICAL STRATEGIES FOR RESOLVING CONFLICT 327 (2d ed. 1996)); *see also* Astor, *supra* note 11, at 74 ("Significant attacks on mediator neutrality have come from academics who have pointed out, trenchantly and repeatedly, that mediators are not neutral. Research has clearly demonstrated that mediators do inject their own values into mediation.").

166. Mulcahy, *supra* note 162, at 511.

167. *Id.* at 516.

168. *Id.* at 516–17.

reactions were common.[169] Their mediation training "assumed that they could keep such negative evaluations of the disputants at bay."[170] However, the mediators felt constrained by an expectation of neutrality, as the expectation "was impossible to achieve" and "made them feel as though they were constantly doomed to failure."[171]

A second study showed that mediators influence the content and outcome of mediations by instigating party engagement at certain times in the process to make certain outcomes more likely.[172] This study looked at divorce mediations, analyzing data from forty-five mediation sessions which covered fifteen cases handled by three mediators.[173] Researchers found that mediators directed the process towards the outcomes they favored.[174] "The pressure that the mediator exerts toward the favored and against the disfavored outcome is largely managed by differentially creating opportunities to talk through the favored option rather than, for example, repeatedly producing evaluative statements about the positions of the two clients or the options open to them."[175] The authors label this technique "selective facilitation"[176] and admonish that it should be "introduced with sufficient clarity for clients to be able to recognize it and choose whether to go along with it."[177]

An additional layer should be explored to address concerns of partiality in actual mediator behavior: the danger of unconscious bias against a party. As previously described, research shows the

---

169. *Id.* at 517.
170. *Id.*
171. *Id.*
172. David Greatbatch & Robert Dingwall, *Selective Facilitation: Some Preliminary Observations on a Strategy Used by Divorce Mediators*, 23 LAW & SOC'Y REV. 613 (1989).
173. *Id.* at 617.
174. *Id.* at 618. Information from the sessions "demonstrates that the mediator is working with notions of what kind of settlement would be desirable (a favored outcome) and what kind of settlement would be undesirable (a disfavored outcome), and seeks to guide the interaction accordingly." *Id.*
175. *Id.* at 636. "More commonly, mediators seem to proceed not by using the negative power of a veto but through the positive power of encouraging discussion in specific directions." *Id.* at 617.
176. *Id.* at 618.
177. *Id.* at 639. "Mediator influence becomes a problem only when formal and substantive neutrality are confused so that the pressure becomes invisible or when the choice of goals remains a purely personal matter rather than one for which the practitioner may be socially accountable." *Id.*

influence of implicit bias on our evaluation of others, judgments, and behavior, which is often inconsistent with express statements. "[E]x ante exhortation not to be intentionally unfair will do little to counter implicit cognitive processes, which take place outside our awareness yet influence our behavior."[178] In their introductory comments to the parties, mediators generally state that they will act in a neutral and impartial manner. Ethical and professional standards impose on mediators a moral imperative to avoid discrimination in their mediations. It is up to the parties to prove discriminatory treatment, even though people often do not perceive discrimination. "A behavioral realist analysis has demonstrated that such a model of explicit discrimination is not up to the task of responding to implicit bias, which is pervasive but diffuse, consequential but unintended, ubiquitous but invisible."[179]

Decades ago, critics cautioned that the mediation process may be particularly ill-suited to identify and confront discriminatory behavior.[180] As Richard Delgado and his colleagues warned, "ADR might foster racial or ethnic bias in dispute resolution."[181] Because formal adjudication explicitly manifests "societal norms of fairness and even-handedness" through symbols (flag, black robe), ritual, and rules, the adversarial process counteracts bias among legal decision makers and disputants.[182] These commentators conclude that members of the majority are most likely to show prejudicial behavior in informal ADR settings.[183] They argue that

> ADR is most apt to incorporate prejudice when a person of low status and power confronts a person or institution of high status and power. In such situations, the party of high status is more likely than in other situations to attempt to call up prejudiced responses; at the same time, the individual of low status is less

---

178. Kang & Banaji, *supra* note 140, at 1079.
179. *Id.* at 1079–80 (citing Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 CALIF. L. REV. 1 (2006)) ("Recognition of the pervasiveness of implicit bias lends support to a structural approach to antidiscrimination law.").
180. Richard Delgado et al., *Fairness and Formality: Minimizing the Risk of Prejudice in Alternative Dispute Resolution*, 1985 WIS. L. REV. 1359.
181. *Id.* at 1367.
182. *Id.* at 1387–88.
183. *Id.* at 1391.

likely to press his or her claim energetically. The dangers increase when the mediator or other third party is a member of the superior group or class.[184]

To test the "informality hypothesis" that the effects of gender and ethnicity will be greater in mediated rather than adjudicated small claims cases, Gary LaFree and Christine Rack examined ethnicity and gender among participants and mediators in Bernalillo County, New Mexico ("MetroCourt study").[185] These researchers compared the impact of disputants' ethnicity and gender on monetary outcomes in 312 adjudicated and 154 mediated civil cases.[186] They found support for the informality hypothesis (i.e., disparities between Anglo males and others will be particularly significant in mediation) in contrasts between minority and Anglo claimants.[187] "The strongest support for the informality hypothesis is for minority male claimants, who received significantly lower MORs [monetary outcome ratios] in mediation, even when case variables are controlled for."[188] The study found no evidence that minorities or women were "especially disadvantaged as respondents in mediation."[189] The researchers concluded there was some support for an informality hypothesis, i.e., "that ethnic and gender disparities are greater in mediation than in adjudication."[190]

LaFree and Rack also sought to test the "disparity hypothesis" that minority and female disputants will achieve less favorable outcomes than majority and male parties whether their cases are adjudicated or mediated, and they found "considerable support" for it.[191] Data for mediated outcomes showed that minority men and women received significantly lower MORs as claimants, and minority men paid

---

184. *Id.* at 1402–03. For a response to Delgado's criticisms, see Sara Kristine Trenary, *Rethinking Neutrality: Race and ADR*, 54 DISP. RESOL. J. 40, 44 (1999).

185. *See generally* Gary LaFree & Christine Rack, *The Effects of Participants' Ethnicity and Gender on Monetary Outcomes in Mediated and Adjudicated Civil Cases*, 30 LAW & SOC'Y REV. 767 (1996).

186. *Id.* at 771.

187. *Id.* at 778.

188. *Id.* at 780.

189. *Id.* at 778.

190. *Id.* at 789.

191. *Id.* at 788.

significantly more as respondents.[192] The study's overall results showed

> the strongest evidence of ethnic and gender disparity in the treatment of minority claimants in mediation. In the analysis including product terms, both minority male and female claimants received significantly lower MORs – even when we included the nine case-specific and repeat-player variables. Of greatest concern is the fact that this disparity was only present in cases mediated by at least one Anglo mediator. Cases mediated by two minorities resulted in lower MORs, regardless of claimant ethnicity.[193]

Rack conducted a second MetroCourt study involving a full data set of 603 small claims cases, of which 323 were adjudicated and 280 were mediated.[194] The study looked at a subset of 138 mediated cases which resulted in monetary agreements.[195] Rack compared party negotiations before the mediation with negotiation movement during the session to assess how the mediation process itself affected disputants.[196] She organized data to view cases as status relationships between claimants and the respondents, using five status dimensions: race-ethnicity, gender, socio-economic, corporate, and legal representation.[197] She found that ethnic minority claimants settled for less than Anglo claimants in mediation.[198] Compared to Anglo counterparts, minority respondents admitted higher liability at the outset and reported similar pre-mediation concessions; however, during the mediation sessions minority respondents conceded proportionally more than Anglo respondents to Anglo claimants.[199]

---

192. *Id.* at 780.
193. *Id.* at 789.
194. Christine Rack, *Negotiated Justice: Gender & Ethnic Minority Bargaining Patterns in the MetroCourt Study*, 20 HAMLINE J. PUB. L. & POL'Y 211, 212 (1999).
195. *Id.*
196. *Id.*
197. *Id.*
198. *Id.* at 217. In the total sample, those coded as "minority claimants" were: 182 Hispanics (30.4%), 11 African-Americans (1.8%), 4 Asians (0.7%), 7 Native Americans (1.2%), and 5 "others" (0.8%). Those coded as minority respondents were: 216 Hispanics (36.1%), 22 African-Americans (3.7%), 11 Asians (1.8%), 5 Native Americans (0.8%), and 14 "others" (2.3%). *Id.* at 238.
199. *Id.* at 249.

"In sum, patterns shown here reflected firm bargaining by higher structural status claimants (high initial demands, concession resistance, undermatching, and little end stage concession-making). At the opposite pole, minority claimants were the softest bargainers."[200] Interestingly, "claimant ethnicity was the significant factor differentiating respondent concession-making; Anglos and men were more willing to pay Anglo than minority claimants."[201] According to Rack, the study showed that "Anglos and women [are] more likely to show insider bias."[202]

Mediators in Rack's study exhibited "Anglo-protective bias."[203] "Especially when the respondent was Anglo, mediators' status deference and ethic of 'neutrality' became a means through which the mediation environment served to support exploitation of soft bargaining."[204] Rack observed that "[o]vert prejudice was rarely acknowledged by disputants or recognized by mediators although the effects were apparent in the outcomes."[205] Noting that "[n]on-dominant groups may hold different fairness values, hold unequal power in negotiations with more dominant parties, and accept disadvantaged outcomes," Rack concluded that "those who are traditionally perceived as less competent continue to be perceived that way persistently so that hierarchies are recreated through a process of self-fulfilling prophecy. Attempts to break free of others' expectations are often negatively misperceived and actively discouraged until less privileged actors retreat from trying."[206]

Rack's MetroCourt study raises concerns that "insider bias" and "Anglo-protective" behavior on the part of mediators, along with settlement pressure to avoid perceived risks of adjudication, put minority parties at a significant disadvantage. Her case studies "suggest what appeared to be primary mediator patterns in these cases; Anglo mediators leaned on external status characteristics to

---

200. *Id.* at 253.
201. *Id.* at 258.
202. *Id.* at 289.
203. *Id.* at 273.
204. *Id.* at 262.
205. *Id.* at 276.
206. *Id.* at 230–31 (citing Cecilia L. Ridgeway, *Interaction and the Conservation of Gender Inequality: Considering Employment*, 62 AM. SOC. REV. 218, 218–35 (1997)).

grant legitimacy in the absence of cultural understanding, a pattern that apparently reinforced a pattern of hierarchy acceptance within the minority culture."[207] Rack noted, "The interest-based negotiation process and the mediators' often unexamined and unintended influence (or lack thereof), offered various opportunities for betrayals of justice. . . . Minority disputants, not Anglo women, manifested bargaining patterns that implied socialization patterns that could be and were substantively exploited by more dominant parties."[208] Rack concluded that "data suggested that the most imbalanced outcomes resulted from settlement pressure through constructing non-monetary substitutes for monetary claims, and by invoking, perhaps misrepresenting, evidentiary rules to discourage disputants from adjudication."[209]

Unique conditions of the mediation process may contribute to discriminatory mediator action (or inaction) in another way. In *Race as Proxy: Situational Racism and Self-Fulfilling Stereotypes*, Lu-in Wang examines the influence of situational context on discriminatory behavior in social interactions.[210] Wang argues that race functions as a proxy for negative characteristics associated with skin color, such as "laziness, incompetence, and hostility . . . lack of patriotism or disloyalty to the United States . . . susceptibility to some diseases . . . [and] criminality and deviance."[211] Wang contends that "fewer individuals than in the past are likely to be motivated by discriminatory animus. . . . Most of us are afflicted instead with unconscious cognitive and motivational biases that lead us to reflexively categorize, perceive, interpret the behavior of, remember, and interact with people of different races differently."[212]

---

207. *Id.* at 263. The minorities involved were Latinos. Rack expressly stated that the same patterns may not be found in research with other minority groups. *Id.*

208. *Id.* at 294–95. "Disparate outcomes were created by apparently soft bargaining that was leveraged by mediators and exploited by opportunistic respondents into greater concessions. Minority claimants were vulnerable to suggestions that they could not expect much from their judicial alternative." *Id.* at 286.

209. *Id.* at 296.

210. Lu-in Wang, *Race as Proxy: Situational Racism and Self-Fulfilling Stereotypes*, 53 DEPAUL L. REV. 1013 (2004).

211. *Id.* at 1013–14. Proxy captures the unconscious and habitual "'default' manner in which race often influences decision-making." *Id.* at 1015.

212. *Id.* at 1017.

Wang advocates an examination of "social constraints" as powerful unseen influences on discriminatory behavior.[213] Contextual circumstances and "external factors" work to create "channel factors" which direct behavior by (1) determining how an individual defines a situation, and (2) channeling her behavior by indicating the appropriate conduct for that situation, "essentially opening or closing pathways for action."[214] Wang cites studies that show that "situations that include clear indications of right and wrong behavior [] tend to lessen the likelihood of discrimination."[215] Normative ambiguity tends to promote discrimination and "the power of ambiguity to channel discrimination goes hand-in-hand with its ability to mask it."[216] Normative ambiguity can arise where appropriate behavior in a particular context is not clearly identified and where clearly negative behavior can be justified on a basis other than race.[217] Stated another way, "normative clarity discouraged racial bias, but normative ambiguity channeled it."[218]

Could normative ambiguity in the mediation process channel biased mediator behavior as Wang posits? Mediators lack the surety of clearly defined rules of intervention. Among mediation professionals, there is little normative consensus regarding appropriate actions and behavior. The mediator's judgments about the parties, her decision to intervene or remain passive at any given time, and her use of various techniques to encourage agreement may be rationalized as "neutral," thus masking bias. An individual "is likely to discriminate in ambiguous situations despite her egalitarian values and lack of prejudice, because she may not be aware of the need to monitor her response and because racial stereotypes are

---

213. *Id.* at 1025.
214. *Id.* at 1026 (citing LEE ROSS & RICHARD E. NISBETT, THE PERSON AND THE SITUATION: PERSPECTIVES OF SOCIAL PSYCHOLOGY 10 (1991)).
215. *Id.* at 1038.
216. *Id.*
217. *Id.* at 1038–39. Citing juror studies, Wang notes that subjects were more likely to engage in discriminatory behavior when they could point to a non-discriminatory reason to rationalize their actions. For example, subjects might rationalize that verdicts were motivated by a desire to not let a guilty person go free rather than by racial bias. *Id.* at 1043.
218. *Id.* at 1039.

always accessible and automatically activated, and will lead her to discriminate despite her best intentions."[219]

Against this backdrop of implicit bias research and the operation of mediator partiality in actual practice, the next Part returns to the case scenario as a vehicle to contemplate subtle dynamics that might operate within a discrete mediation context.

IV. APPLICATION TO ASIAN AMERICANS[220] IN MEDIATION

Turning back to the Michigan small claims mediation described in the Introduction, I hope to stimulate a fresh inquiry into mediator actions. What influence, if any, might implicit bias have had on the mediators' perception and judgment of the parties? Is it possible that the mediators unintentionally favored the business owner in the mediation? As in the MetroCourt study, did the mediators demonstrate "insider bias" or in-group protectionism? Could the mediators' attitudes toward the homeowners have been colored by Asian stereotypes? In what ways could unconsciously held stereotypic views of a group operate in a seemingly simple non-racialized dispute? "[S]tereotypes about ethnic groups appear as part of the social heritage of society. They are transmitted across generations as a component of the accumulated knowledge of a society. They are as true as tradition, and as pervasive as folklore. No person can grow up in a society without having learned the stereotypes assigned to the major ethnic groups."[221] At the outset, let me state that I believe the mediators conducted the process earnestly and without indication of explicit negative or positive attitudes toward either party. They showed no outright bias, favoritism, or prejudice during the mediation. They employed a facilitative style of

---

219. *Id.* at 1045 (citing Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 15–16 (1989)).

220. The United States Census Bureau defines Asian-American as "[a] person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. It includes 'Asian Indian,' 'Chinese,' 'Filipino,' 'Korean,' 'Japanese,' 'Vietnamese,' and 'Other Asian.'" U.S. Census Bureau, *State & County QuickFacts*, CENSUS.GOV, http://quickfacts.census.gov/qfd/meta/long_RHI425200.htm (last visited Nov. 30, 2010).

221. HOWARD J. EHRLICH, THE SOCIAL PSYCHOLOGY OF PREJUDICE 35 (1973).

mediation as taught in the required forty-hour Michigan Civil Mediation Training.[222] I suggest that the likelihood that implicit bias operated is as great as, or even greater than, the likelihood it did not.

A. Evolution of Asian American Stereotypes

Asian American stereotypes have notably evolved over the past century. Chinese in the United States in the late 1800s were characterized as opium-smoking, morally deficient sub-humans.[223] Fearing the "yellow peril" at the turn of the nineteenth century, Americans portrayed Chinese as military, cultural, or economic enemies and unfair competitors.[224] Courts and legislatures have a long history of discrimination against Asian Americans.[225] In *People v. Hall*,[226] Chinese were described as people

> whose mendacity is proverbial; a race of people whom nature has marked as inferior, and who are incapable of progress or intellectual development beyond a certain point, as their history has shown; differing in language, opinions, color, and

---

222. This assumes the training they underwent was similar to the one I completed in order to mediate small claims cases.

223. Neil Gotanda, *Asian American Rights and the "Miss Saigon Syndrome," in* ASIAN AMERICANS AND THE SUPREME COURT: A DOCUMENTARY HISTORY 1087 (Hyung-Chan Kim ed., 1992); RONALD T. TAKAKI, STRANGERS FROM A DIFFERENT SHORE: A HISTORY OF ASIAN AMERICANS 99–112 (rev. ed. 1998); Keith Aoki, *"Foreign-ness" & Asian American Identities: Yellowface, World War II Propaganda, and Bifurcated Racial Stereotypes*, 4 ASIAN PAC. AM. L.J. 1, 18–23 (1996); Pat K. Chew, *Asian Americans: The "Reticent" Minority and Their Paradoxes*, 36 WM. & MARY L. REV. 1, 12–15 (1994).

224. TAKAKI, *supra* note 223, at 81; *see also* Natsu Taylor Saito, *Model Minority, Yellow Peril: Functions of "Foreignness" in the Construction of Asian American Legal Identity*, 4 ASIAN L.J. 71, 72 (1997).

225. For example, the Chinese Exclusion Act of 1882, ch. 126, 22 Stat. 58, barred Chinese immigration and "caused untold suffering and hardship, separating families, creating a society of single men, and institutionalizing hostility, prejudice against and isolation of Chinese immigrants and Chinese Americans." City & Cnty. of S.F. Bd. Res. 363–09 (San Francisco, Cal. Sept. 15, 2009). Resolution No. 363-09 of the San Francisco Board of Supervisors "acknowledg[es] the regrettable role that San Francisco has played in advancing the policies of the Chinese Exclusion Act of 1882, the first federal law to discriminate against a specific group solely on the basis of race or nationality." *Id.*

226. 4 Cal. 399 (1854).

physical conformation; between whom and ourselves nature has placed an impassable difference.[227]

The Supreme Court upheld the denial of citizenship to Japanese and Hindus from India, concluding that the forefathers intended to exclude "Asiatics" from naturalization and citizenship.[228] "Alien Land Laws" denied Americans of Japanese ancestry the right to own property.[229] Fervent anti-Japanese sentiment and suspicion ultimately led to the incarceration of 120,000 Japanese American citizens and legal permanent residents during World War II.[230]

The next forty years witnessed a shift in the way Asian Americans were perceived. As time passed, Asian Americans went from being a "bad" minority to a "good" minority. They were viewed as smart, industrious, and unassuming.[231] William Peterson first coined the term "model minority" in a 1966 *New York Times Magazine* article about Japanese Americans.[232] Asian Americans were held up as examples of minority success through hard work, sacrifice, following rules, keeping their noses to the grindstone, and minding their own business. Asian Americans, in short, achieved the American Dream. Americans have embraced the model minority perception as the contemporary Asian American stereotype.[233]

---

227. *Id.* at 405. The court found that section 13 of the Act of April 16, 1850, prohibited Chinese people from testifying in favor of or against white men. *Id.* The court thus reversed the conviction of a white man who was found guilty of murder based on the testimony of Chinese witnesses. *Id.*

228. Ozawa v. United States, 260 U.S. 178, 195–96 (1922). In *Ozawa*, the Court found that section 2169 of the Revised Statutes, which limited naturalization to aliens who were "free white persons" and to aliens of African descent, applied to the Naturalization Act of June 29, 1906, ch. 3592, secs. 355–353, § 1, 34 Stat. 596 (1906). *Ozawa*, 260 U.S. at 194. This made the Japanese appellant ineligible for naturalization because he was not a free white person. *Id.* at 198; *see also* United States v. Thind, 261 U.S. 204 (1923) (determining that the term "free white persons" was to be interpreted as a common man would understand it; that the term was found to be synonymous with the word "Caucasian"; and that a high caste Hindu of full Indian blood was not included in that term).

229. Keith Aoki, *No Right to Own?: The Early Twentieth-Century "Alien Land Laws" as a Prelude to Internment*, 40 B.C. L. REV. 37, 38 (1998).

230. ERIC K. YAMAMOTO ET AL., RACE, RIGHTS AND REPARATION: LAW AND THE JAPANESE AMERICAN INTERNMENT 4 (2001).

231. Saito, *supra* note 224, at 71.

232. Chew, *supra* note 223, at 24 (citing William Petersen, *Success Story, Japanese-American Style*, N.Y. TIMES MAG., Jan. 9, 1966, at 20–21, 33, 36, 40–41, 43).

233. *Id.* at 24.

The model minority stereotype, like all stereotypes, is inaccurate. Lumping all Americans of Asian descent into one homogeneous category ignores vast differences among the many ethnicities. Dozens of different ethnic groups fall under the "Asian American" umbrella.[234] In fact, the pan-Asian identity reflected in the term did not develop until the 1960s.[235] Three main factors complicate any assumption of Asian Americans as a monolithic group: country of ancestry, length of residence in the United States, and gender.[236]

The model minority myth also has a negative side. Quiet, high achieving, workaholic go-getters may also be seen as cut-throat, inscrutable, and sneaky.[237] Asian Americans are viewed as skilled in scientific, technical, and quantitative fields, but lacking in verbal, social, and interpersonal skills.[238] This positive/negative duality of the stereotype is "akin to the paradoxical topology of a mobieus strip. If pressed, the so-called 'good' attributes . . . easily transform into the 'bad' attributes . . . and vice versa."[239]

The model minority myth masks challenges faced by Asian Americans who are over-credited with ascension on the ladder of success. The poverty rate for Asian Americans is almost twice that of white Americans.[240] Family income comparisons fail to recognize that Asian families typically have more workers per family than families with higher individual incomes.[241] Perceptions of Asian

---

234. *Id.* at 25.
235. YAMAMOTO ET AL., *supra* note 230, at 269–70.
236. Chew, *supra* note 223, at 26. For example, a fourth-generation Japanese American in California has very little in common with a recent Hmong immigrant in Minnesota, and Native Hawaiians have a vastly different set of experiences and perspectives than mainland Asian Americans.
237. Saito, *supra* note 224, at 72; Chew, *supra* note 223, at 38.
238. The "Asians are good at math" stereotype is so strong that it is even internalized by Asian Americans. The Math Test study by Margaret Shih showed that by unconsciously activating a particular identity (Asian) in Asian American female undergraduates, performance on a difficult math test was improved. Conversely, when female identity was unconsciously activated, the students' performance was depressed downward. Margaret Shih et al., *Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance*, 10 PSYCHOL. SCI. 80 (1999).
239. Aoki, *supra* note 223, at 35–36.
240. Saito, *supra* note 224, at 90 (citing William R. Tamayo, *When the "Coloreds" Are Neither Black Nor Citizens: The United States Civil Rights Movement and Global Migration*, 2 ASIAN L.J. 1, 15 n.97 (1995)).
241. TAKAKI, *supra* note 223, at 475.

Americans include the belief that they are not the targets of racial discrimination[242] and that they are represented throughout the ranks of industries and professions.[243] Discussing Asian Americans, one scholar commented that "[a]lthough they are often needy and disadvantaged, they are not perceived as facing any obvious barriers greater than those of previous immigrant groups. . . . For example, there is less concern about [them] than about blacks, and they are less negatively stereotyped."[244] The model minority myth sends a message that Asian American claims of discrimination are not to be taken seriously.[245]

The stereotype that Asian Americans are deferential and unassertive hurts their potential to advance in various professional fields. Asian Americans are under-represented at the top levels of corporate, legal, and commercial management.[246] "[B]eliefs about Asian Americans as individually passive, obedient, hardworking, and socially inept encourage employers to hire them, but not promote them to upper levels of management. The combined effect of these racial beliefs produces a glass ceiling."[247] Stereotyping of this nature is evident in a recent case involving the exclusion of Asian Americans as grand jury forepersons.[248] In *Chin v. Runnels*, a

---

242. Gotanda, *supra* note 223, at 1091. One study found that nearly 40 percent of whites thought that with regard to job and housing discrimination, Asian Americans experience "little" or "none." Chew, *supra* note 223, at 8 (citing Michael McQueen, *Voters' Responses to Poll Discloses Huge Chasm Between Social Attitudes of Blacks and Whites*, WALL ST. J., May 17, 1991, at A16). In contrast, another study indicated that 49 percent of Asian Americans stated they had experienced discrimination. *Id.* at 8 (citing *Study Says Asians Feel Bias More Than Hispanics*, L.A. DAILY J., Dec. 12, 1985, at 1).

243. Chew, *supra* note 223, at 46.

244. David O. Sears, *Racism and Politics in the United States*, *in* CONFRONTING RACISM: THE PROBLEM AND THE RESPONSE 76, 95 (Jennifer L. Eberhardt & Susan T. Fiske eds., 1998).

245. Gotanda, *supra* note 223, at 1089.

246. Chew, *supra* note 223, at 47–49.

247. Don Operario & Susan T. Fiske, *Racism Equals Power Plus Prejudice: A Social Psychological Equation for Racial Oppression*, *in* CONFRONTING RACISM: THE PROBLEM AND THE RESPONSE 33, 52 (Jennifer L. Eberhardt & Susan T. Fiske eds., 1998).

248. *See* Darren Seiji Teshima, *A "Hardy Handshake Sort of Guy": The Model Minority and Implicit Bias About Asian Americans in* Chin v. Runnels, 11 ASIAN PAC. AM. L.J. 122 (2006) (arguing that court officials, implicitly biased because of the model minority stereotype, believed that Asian Americans were not good forepersons because they were not good leaders); *see also* Benson, *supra* note 135, at 47 (hypothesizing that a judge who accepted prejudiced stereotypes of Asian Americans as "introverted and timid" would not select a Chinese American foreperson).

Chinese-American defendant claimed that exclusion of Chinese-Americans, Hispanic-Americans, and Filipino-Americans as grand jury forepersons violated his right to equal protection under the Fourteenth Amendment.[249] Petitioner established a prima facie case of discrimination in the selection of jury forepersons under a process in which the judge and others identified "leadership capabilities."[250] The court expressly entertained the claim that unconscious biases may have contributed to this forty year exclusion, concluding that there may be "a sizeable risk that perceptions and decisions made here may have been affected by unconscious bias."[251]

The second pervasive stereotype of Asian Americans is known as the "perpetual foreigner syndrome."[252] This element of "foreignness" is rooted in the racial categorization of Asians as the "Mongolian or yellow race," as distinguished from the "white or Caucasian race."[253] Even Asian Americans who are native-born citizens have historically been viewed as foreigners.[254] Foreignness became linked with political disloyalty.[255] The imprisonment of Japanese Americans, many of whom were U.S. citizens, during World War II presents a glaring example of this conflation of native-born Asian American citizens with a foreign enemy.[256] Similarly, the foreignness-disloyalty connection has been applied to Korean Americans and Vietnamese Americans during conflicts with Asian countries.[257] The imagery of

249. Chin v. Runnels, 343 F. Supp. 2d 891, 892 (N.D. Cal. 2004).
250. *Id.* at 896–97, 901. Statistical evidence showed that between 1960 and 1996, not one Chinese American, Filipino American, or Hispanic American served as jury foreperson, and that the statistical likelihood of this occurring was 0.0003%. *Id.* at 895.
251. *Id.* at 908. The court denied petitioner's habeas claim but intimated that under de novo review, petitioner likely would have been granted relief. *Id.* at 905–08.
252. FRANK H. WU, YELLOW: RACE IN AMERICA BEYOND BLACK AND WHITE 79–129 (2002); Saito, *supra* note 224, at 76; Gotanda, *supra* note 223, at 1097; Chew, *supra* note 223, at 34.
253. *See* Saito, *supra* note 224, at 78 (citing *In re* Ah Yup, 1 F. Cas. 223 (D. Cal. 1878)); *see also* Aoki, *supra* note 223, at 9–10.
254. Saito, *supra* note 224, at 75–76; *see also* Chew, *supra* note 223, at 35.
255. Saito, *supra* note 224, at 82.
256. YAMAMOTO, *supra* note 230, at 4; Saito, *supra* note 224, at 81–83. General John L. DeWitt, leader of the Western Defense Command who favored internment of West Coast Japanese Americans, famously said, "A Jap's a Jap. . . . It makes no difference whether he is an American citizen, [theoretically,] he is still a Japanese." YAMAMOTO ET AL., *supra* note 230, at 99.
257. Saito, *supra* note 224, at 84.

Asian   Americans   as   the   enemy   persists   through   economic
competition and American trade protectionism, from the 1980s
"Japan bashing" caused by automotive competition to imposition of
tariffs on cheaper tires imported from China in 2009.[258]

Social cognition research by Thierry Devos and Mahzarin Banaji
in 2005 substantiated the perpetual foreigner syndrome. Their study
revealed that Asian Americans are perceived as being less American
than both Whites and African Americans.[259] Experimental subjects
linked American-ness more with white Europeans (e.g., Hugh Grant)
than with famous Asian Americans (e.g., Connie Chung).[260] "The
conclusion that can be drawn on the basis of the six studies presented
here is unambiguous. To be American is to be White."[261] The model
minority myth and perpetual foreigner syndrome were confirmed by
scientific method in 2009. A survey conducted by Harris Interactive
in January 2009 using a computer-assisted telephone interviewing
system ("C100 Survey") assessed current attitudes toward Chinese

---

258.   WU, *supra* note 252, at 70, 88–89; Peter Whoriskey & Anne Kornblut, *U.S. to Impose
Tariff on Tires From China*, WASH. POST, Sept. 12, 2009, at A1.

259.   Thierry Devos & Mahzarin R. Banaji, *American = White?*, 88 J. PERSONALITY & SOC.
PSYCHOL. 447, 463 (2005). Readers may recall MSNBC's gaffe in 1998, running the headline
"American Beats Out Kwan" on a story about Tara Lipinski's defeat of her favored U.S.
teammate, Michelle Kwan. *See* Steve Mirsky, *Birth of a Notion: Implicit Social Cognition and
the 'Birther' Movement*, SCI. AM., Oct. 2009, at 100.

260.   Devos & Banaji, *supra* note 259, at 456–57.

261.   *Id.* at 463. Devos conducted a more recent study that found that the participants more
closely associated Hillary Clinton with American sentiments than they did Barack Obama. This
was true regardless of whether race, gender, or personal identity were emphasized, though it
was more pronounced when race was emphasized. Thierry Devos, Debbie S. Ma & Travis
Gaffud, *Is Barack Obama American Enough to Be the Next President?: The Role of Ethnicity
and National Identity in American Politics*, http://www.rohan.sdsu.edu/~tdevos/thd/Devos_
spsp2008.pdf. The researchers concluded, "A Black candidate is implicitly conceived of as
being less American than a White candidate when perceivers focus on the targets' ethnicity."
Id.; *see also* Gregory S. Parks, Jeffrey J. Rachlinksi & Richard A. Epstein, Debate: *Implicit
Bias and the 2008 Presidential Election: Much Ado about Nothing?*, 157 U. PA. L. REV.
PENNUMBRA 210 (2009), *available at* http://www.pennumbra.com/debates/pdfs/Implicit
Bias.pdf. Parks, Rachlinski, and Epstein argue that while Obama's election represents a
monumental stride forward for race relations, any announcement of a post-racial America is
premature because of the race-tinged aspects of the election, including perceptions of Obama as
insufficiently patriotic or American. Citing implicit bias, they caution that "[m]odern racism no
longer produces an overt smoking gun marking its influence; one has to look fairly carefully to
find its influence. It operates not as an absolute barrier, but as a kind of tax on members of
racial minorities. It facilitates certain negative assumptions through an invisible influence." *Id.*
at 214.

Journal of Law & Policy   [Vol. 34:71

and Asian Americans.[262] The survey covered issues such as "race relations, social equality, immigration, and factors influencing public attitudes."[263] It compared responses from the general population sample and responses from the Chinese American sample. Related to the model minority myth, "[o]ver half of both the general population and Chinese Americans believe Asian Americans achieve a higher degree of overall success often or always in comparison to other Americans."[264] Reflecting perpetual foreigner status, 74 percent of the general population sample overestimated the proportion of the U.S. population that is made up of Asian Americans; contemporaneously, 51 percent underestimated the population of Asians born in the United States.[265] Judging loyalty, three-quarters of the Chinese American over-sample said that Chinese Americans "would support the U.S. in military or economic conflicts between the U.S. and China," but only about half of the general population "believe Chinese Americans would support the U.S. in such conflicts."[266] On racial profiling, only two-fifths of the general population think the FBI might prematurely arrest an Asian American;[267] more than half of the Chinese American respondents believe the FBI would arrest an Asian American without sufficient evidence.[268]

---

262. COMMITTEE OF 100 & HARRIS INTERACTIVE, STILL THE "OTHER?": PUBLIC ATTITUDES TOWARD CHINESE AND ASIAN AMERICANS (2009), *available at* http://www.survey.committee100.org/2009/files/FullReportfinal.pdf. The survey followed up on a 2001 study "to gauge shifts in attitudes" and to "explore factors that help formulate perceptions and the reasoning behind attitude changes." *Id.* at 8. The survey used "split samples to compare attitudes toward Chinese Americans, Asian Americans, and other racial or religious groups. In addition to the general population sample, an over-sample of Chinese Americans was conducted." *Id.*
263. *Id.*
264. *Id.* at 42.
265. *Id.* at 40.
266. *Id.* at 43.
267. *Id.* at 45.
268. *Id.* at 44. For a discussion of the Wen Ho Lee case as a recent example of Asian American racial profiling, see Neil Gotanda, *Comparative Racialization: Racial Profiling and the Case of Wen Ho Lee*, 47 UCLA L. REV. 1689, 1692–94 (2000).

2010]   Implicit Bias and the Illusion of Mediator Neutrality   117

B. Revisiting the Small Claims Case

Returning to the small claims mediation, let us reexamine the mediators' conclusion that the door was closed. Presumably, the mediator team was aware of the importance of mediator neutrality to their role and to the sustention of a legitimate process.[269] The Michigan Standards of Conduct for Mediators require the mediators to "remain impartial."[270] Studies find that implicit bias is so pervasive, it is likely most of us are affected.[271] Also, IAT data show unconscious racial bias among European American test takers toward disadvantaged groups.[272] Dissociation between implicit and explicit attitudes is common, so these mediators may hold explicit anti-discrimination attitudes and espouse egalitarian views but still have implicit racial biases.[273]

> At a very early age, young Americans learn the stereotypes associated with the various major social groups. These stereotypes generally have a long history of repeated activation, and are apt to be highly accessible, whether or not they are believed. . . . [O]ne can be "nonprejudiced" as a matter of conscious belief and yet remain vulnerable to the subtle cognitive and behavioral effects of implicit stereotypes.[274]

Also, implicit attitudes are better predictors of some behaviors than explicit attitudes.[275] It is conceivable that the mediators interacted

---

269. With regard to impartiality, the Standards of Conduct for Mediators put forward by the State Court Administrative Office of the Michigan Supreme Court state:

> A mediator shall conduct the mediation in an impartial manner. The concept of mediator impartiality is central to the mediation process. A mediator shall mediate only those matters in which it is possible to remain impartial and even-handed. If at any time the mediator is unable to conduct the process in an impartial manner, the mediator is obligated to withdraw.

STANDARDS OF CONDUCT FOR MEDIATORS (State Court Admin. Office, Mich. Supreme Court 2001), *available at* http://www.courts.michigan.gov/scao/resources/standards/ odr/conduct.pdf.
    270. *Id.*
    271. Lane, Kang & Banaji, *supra* note 71, at 433–37.
    272. Greenwald & Krieger, *supra* note 66, at 955–58.
    273. *Id.* at 955–56.
    274. Krieger & Fiske, *supra* note 133, at 1033.
    275. Lane, Kang & Banaji, *supra* note 71, at 435–37.

with the parties in a way that was unconsciously more favorable toward the business owner and less favorable toward the homeowners. We learned that group membership implicitly affects a person's identity formation and unconscious expressions of feeling and thought, and that in-group favoritism is strong.[276] "A person may have a view of herself as egalitarian but find herself unable to control prejudicial thoughts about members of a group, perhaps including groups of which she is a member."[277] A person's membership in a group implicitly affects that person's identity formation and "ingroup bias occurs automatically or unconsciously under minimal conditions."[278]

Considering potential mediator bias and favoritism in light of the science of implicit social cognition, it is conceivable that Asian American stereotypes were automatically activated when the mediators met the homeowners. "[M]erely encountering a member of a stereotyped group primes the trait constructs associated with and, in a sense, constituting, the stereotype. Once activated, these constructs can function as implicit expectancies, spontaneously shaping the perceiver's perception, characterization, memory, and judgment of the stereotyped target."[279] Clearly, race alters interpersonal, intrapersonal, and intergroup interactions.[280]

With activation of the stereotype that Asians are untrustworthy, the mediators may have unconsciously viewed the homeowners as less credible or as giving a less reliable account of the rug cleaning situation. They may have implicitly favored the story put forward by the carpet cleaner (in-group) and discredited the version offered by the homeowners (devalued out-group). Perceiving the homeowners as

---

276. Thierry Devos & Mahzarin R. Banaji, *Implicit Self and Identity*, *in* HANDBOOK OF SELF AND IDENTITY 153, 154–58 (M.R. Leary & J.P. Tangney eds., 2003), *reprinted in* 1001 ANNALS N.Y. ACAD. SCI. 177, 179–85 (2003).
277. *Id.* at 179.
278. *Id.* at 185.
279. Krieger & Fiske, *supra* note 133, at 1033.
280. Kang, *supra* note 136, at 1493; *see also* Kang & Banaji, *supra* note 140, at 1085 ("An individual (target) is mapped into a social category in accordance with prevailing legal and cultural mapping rules. Once mapped, the category activates various meanings, which include cognitive and affective associations that may be partly hard-wired but are mostly culturally-conditioned. These activated meanings then alter interaction between perceiver and target. These [racial] mechanics occur automatically, without effort or conscious awareness on the part of the perceiver.").

foreign may have activated mental links associating them as an "enemy." The mediators may have unconsciously judged the homeowners as less deserving of relief because of the model minority myth and their "success" in relation to the carpet cleaner.

Mediator memory may have played a role here. Experiments reveal a causal relationship between unconscious stereotypes and biases in perception and memory.[281] Memory errors may occur "because of the human mind's heavy reliance on stereotypes during the encoding and recall of information."[282] Justin Levinson conducted a study testing the effect of implicit racial bias on juror memory.[283] After reading a story about an incident (a fight or employment termination) and performing a distraction task, 153 students of diverse backgrounds[284] answered a questionnaire about the story. The race of the actors in the story was a variable (black/white/Hawaiian).[285] Overall, participants misremembered information in a racially biased way against blacks, less so for Hawaiians.[286] Participants recalled aggressiveness of blacks more easily and generated false memories of their aggression, whereas false memory toward the white actor was positive (receiving an award).[287] Recall is more accurate and false memory generation occurs more with stereotype-consistent information.[288] In addition, "cognitive confirmation effect" has been verified experimentally.[289] Once a social schema (e.g., race, gender) has been activated, a person will often actively search for information that supports that schema

---

281. *See* Banaji & Greenwald, *supra* note 94.
282. Justin D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L.J. 345, 376 (2007).
283. *Id.* at 345.
284. *Id.* at 390–91. The study consisted of 71.2 percent women. Approximately 20 percent of the participants were Japanese American, 20 percent were white, 50 percent were of mixed ethnicity, 2 percent were Hawaiian, 4 percent identified as Other, and there were no African Americans. *Id.*
285. *Id.* at 394.
286. *Id.* at 398.
287. *Id.* at 398–99.
288. *Id.* at 400–01.
289. Page, *supra* note 149, at 216–17 (citing John M. Darley & Paget H. Gross, *A Hypothesis-Confirming Bias in Labeling Effects*, 44 J. PERSONALITY & SOC. PSYCHOL. 20, 20 (1983)).

rather than information that is inconsistent, a process that occurs unconsciously.[290]

Discrimination on the basis of the Asian homeowners' accent is another possible influence on the mediators. Mari Matsuda cautions that "discrimination against accent is the functional equivalent of discrimination against foreign origin."[291] Accent discrimination is triggered by "the collective xenophobic unconscious" bias that operates when a different voice is devalued.[292] A prejudiced listener will attach "a cultural meaning, typically a racist cultural meaning, to the accent."[293] Matsuda suggests that awareness that accent discrimination is a potential problem can help listeners avoid unconscious negative reaction to the accents.[294] Interestingly, not all accents evoke negative reactions. Writing about university tenure decisions, an academic observed that accent is usually a factor in tenure decisions when the professor is a member of an Asian, Indian, African, or Middle Eastern culture; it rarely arises in the case of native speakers of European languages.[295] In the Michigan case, the homeowners' accents, coupled with negative Asian stereotypes, may have caused the mediators to devalue their statements which contradicted the carpet cleaner.

---

290. *Id.*
291. Mari J. Matsuda, *Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction*, 100 YALE L.J. 1329, 1349 (1991). Observing that speech can position people socially, Matsuda claims that "certain dialects and accents are associated with wealth and power. Others are low-status with negative associations." *Id.* at 1352 (citing Marc Fisher, *At GWU, Accent is on English for Foreign Instructors: Student Complaints About Teaching Assistants Lead to Testing Program*, WASH. POST, Nov. 29, 1986, at B1); *see also* Beatrice Bich-Dao Nguyen, *Accent Discrimination and the Test of Spoken English: A Call for an Objective Assessment of the Comprehensibility of Nonnative Speakers*, 1 ASIAN L.J. 117, 122 (1994); Kristina D. Curkovic, *Accent and the University: Accent as Pretext for National Origin Discrimination in Tenure Decisions*, 26 J.C. & U.L. 727 (2000); Mary E. Mullin, Comment, *Title VII: Help or Hindrance to the Accent Plaintiff*, 19 W. ST. U. L. REV. 561, 571 (1992); Brant T. Lee, *The Network Economic Effects of Whiteness*, 53 AM. U. L. REV. 1259, 1275 (2004).
292. Matsuda, *supra* note 291, at 1372 (citing ROBERT TAKAKI, FROM DIFFERENT SHORES: PERSPECTIVES ON RACE AND ETHNICITY IN AMERICA (1987)) (asserting that accent discrimination involves "a set of ingrained assumptions that are inevitably lodged in the process of evaluation and in the ways in which we assign values").
293. *Id.* at 1378.
294. *Id.* at 1373.
295. Curkovic, *supra* note 291, at 742–43.

The previous sections of this Article are intended to provoke, not proselytize. The purpose of presenting the small claims scenario is to raise the issue of implicit bias, not to resolve it definitively. As instructors and providers of mediation services, we should understand that mere good intentions to act impartially are insufficient to counter unconscious biases.

> Mediation, despite its image as a neutral procedure in which all values are honored equally and all parties are free to express their points of view, can often be skewed by bias. Mediators often make quick judgments and proffer strong statements infused with their biases, which, though not legally binding, can powerfully impact the outcome of a settlement. . . . Moreover, bias on the part of any mediator can creep into the process in even more subtle ways, such as in the subjective matters of how questioning occurs and how and whether private caucuses are conducted.
>
>   Compounding the problem, it is nearly impossible to accurately observe or address issues of bias in the informal consensus-building environment of mediation, especially because there is an unspoken taboo against acknowledging it.[296]

IV. WHAT DO WE DO?

The prospect of mitigating mediator bias is daunting, but myriad acts and practices within the control of mediators may help address the problem. As a first step, mediation professionals must be realistic and frank about the vast range of mediator behavior and the maneuvers mediators employ to meet the ethical standard of neutrality. We should accept that mediator neutrality is elusive and shape-shifting; it is neither a condition nor characteristic that one possesses or lacks. It is a complex, multi-layered relationship and a system of interaction with the parties that requires constant vigilance. A mediator does not enter a mediation as a "neutral" entity, free from

---

296.   Frederick Hertz, *Bias in Mediation and Arbitration*, CAL. LAW., Nov. 2003, at 37–38.

judgments, values, ideologies, attitudes, and pre-conceived perceptions. Like other human beings, mediators bring prejudices and preferences into the sessions. We should envision neutrality as an unending search, not a state of being.

Realistically, pure impartiality cannot exist in a mediation setting. That said, we should not abandon neutrality as a goal. Rather, mediation practitioners and academics must seek greater understanding and candor about what is done in these confidential, closed-door encounters.

> Neutrality is not an attribute that mediators do, or do not, possess but it is an issue which must be attended to throughout a mediation and which requires constant process of evaluation and decision-making. . . . If we view neutrality through a binary lens, so that it is either present or absent, the research demonstrates as it must, that mediators are not neutral."[297]

As mediators, we should increase our efforts to use the best practices to conduct the process in a way that integrates all aspects of neutrality, i.e., no compromising interests held by the mediator, procedural even-handedness, outcome neutrality, and without bias, prejudice or favoritism toward any party.[298]

To fulfill our commitment to act in a nondiscriminatory manner, it is productive to conceive of mediator neutrality as having both external and internal components.[299] External neutrality consists of conduct and statements to show freedom from bias or favoritism in the way the mediation is conducted. Internal neutrality is the state of being aware of the operation of biases toward the disputants and working to minimize it. I separate bias reduction ideas into these two distinct categories, but I recognize that they coalesce in certain instances. In addition to collecting views from a wide variety of observers, I offer experiences from my law school mediation programs as examples of potentially constructive approaches.

---

297. Astor, *supra* note 11, at 79–80.

298. *See supra* Part I.

299. Rock, *supra* note 50, at 355 ("Internal neutrality refers to the absence of emotions, values or agendas from the mind of the mediator. External neutrality refers to the absence of emotions, values or agendas from the words, actions, and appearance of the mediator.").

A. External Neutrality

The external aspect of neutrality demands paying attention to process attributes, nuances of language and narrative, and the physicality of mediator actions. As practitioners, we are trained to attend to process management and procedures. We strive for external neutrality by conducting an outwardly even process, eliminating conflicts of interest that may arise from proprietary, monetary, relational, and other interests, and abstaining from advocating or pressing for a particular outcome. We seek to ensure our external neutrality through "process policing" techniques: how we engage the parties, manage their interaction, and orchestrate the sessions.[300] A large part of the mediator's job is "maintaining the orderly character of talking and listening, including such matters as organizing the opening and the closing of the session, keeping the parties focused on the current topic, and managing the changes from one topic to another."[301] Management of the agenda goes to the process of interaction, and therefore "can be thought of as being executed in ways that are both formally and substantively neutral."[302]

1. Process Management and Mediator Communication

Mediators manifest external neutrality by being deliberate in planning and conducting each mediation to "place and keep the power of self-determination with the parties, while protecting all parties' abilities to present issues and concerns equally in the mediation session."[303] Practitioners should be mindful of the difference between even-handed process management and "selective facilitation," or maneuvers that are designed to influence and favor certain outcomes. These maneuvers include inhibiting discussion of a

---

300.   External neutrality techniques would include the "*agenda management* that goes on in any orchestrated encounter. . . . Orchestration is one of the means by which speech exchange is ordered in multi-party encounters." Greatbatch & Dingwall, *supra* note 172, at 636 (citation omitted).
   301.   *Id.* at 637.
   302.   *Id.*
   303.   Rock, *supra* note 50, at 356.

disfavored option or moving to close a session without systematic exploration of both parties' preferences.[304]

External neutrality should be assessed through all stages of the process, from pre-mediation preparation through post-mediation evaluation and debriefing. Rather than routinizing procedures for assembly line mediation, mediators should "customize" the sessions for the special dynamics involved.[305] Departure from procedural defaults may be more appropriate under the circumstances. The Michigan mediators followed the general rule for small claims mediation: they asked the party who initiated the matter to make his presentation first. The mediator team may have considered this to be a neutral selection, but it could be perceived as favoring the businessman and disadvantaging the homeowners. After inviting the business owner to speak first, the Michigan mediators posed more inquiries to the business owner than to the homeowners in the joint session and individual sessions. They may have devoted more time to the carpet cleaner and interacted less with the homeowners for various reasons (such as Mr. D's anger, the Ds' accents, or their "foreignness").

External neutrality efforts include consideration of table arrangements and seating arrangements. In the Michigan scenario, the white male mediator sat closer to the business owner. Such an arrangement could create a more intimate conversational dynamic between the two men and give the impression they are "chummy" or in alignment. Both homeowners were seated farther from the mediators than the business owner, making them seem like more remote "outsiders." Both homeowners should have been placed literally "at the table," rather than letting Mrs. D sit behind her husband. If one party is harder to comprehend (perhaps because of accent, soft voice, or looking down), the mediators could alter the arrangement and form a tight circle with no table. Mediators should be careful about chair placement and body positioning so as not to turn their backs toward one disputant more than the other. Special

304. Greatbatch & Dingwall, *supra* note 172, at 637–38.
305. ABA SECTION OF DISPUTE RESOLUTION, TASK FORCE ON IMPROVING MEDIATION QUALITY FINAL REPORT 12–13 (2008), *available at* http://www.abanet.org/dispute/documents/FinalTaskForceMediation.pdf.

challenges may be presented when language interpreters or other third parties are in attendance, as this may make the unassisted party feel outmanned. Interpreters (of American sign language, for example) may need to be seated to accommodate the need to communicate adequately with their clients. Physical limitations of the participants should be considered with external neutrality in mind.[306]

All subtleties of a mediator's mode of communication, including tone of voice, speed of speech, demeanor, eye contact, facial expressions, body language, and physical signals and gestures, are important for attending to external neutrality.[307] Mediators who are fast talkers may disfavor or alienate parties who speak more slowly or who are less fluent in English. We need to be patient with parties who are less articulate or direct than ourselves, and refrain from interrupting, completing sentences, and filling space with words. Regional differences in speech patterns might create mediator affinity with one party over another.[308] Unevenness in eye contact, body placement and movement (sitting forward or leaning back), and attentiveness (looking down while taking notes) may send signals of mediator approval or friendliness, or a lack thereof. When mediating with parties who have physical, cognitive, or intellectual disabilities, we must monitor habits that may inadvertently slight or alienate them. Mediators must be attuned to unintended differential or compensatory treatment (e.g., speaking in a loud voice to a party for whom English is a second language) that may be regarded as treating one participant more positively or negatively than the other. We should be aware of the inadequacy of our usual mannerisms with certain parties; for example, muted visual cues may disadvantage deaf parties who focus more on visual cues and facial expressions.

---

306. For instance, with my limited range of neck motion due to arthritis, as a mediator I must be seated so that I can make eye contact with and view all parties equally.

307. Rock, *supra* note 50, at 358.

308. For example, the East Coast students in my mediation clinic who talk as fast as a "New York minute" often get impatient with parties who speak slowly.

2. Language, Narratives, and Cultural Myths

The importance of language in mediation cannot be overstated. Sarah Burns recommends that mediators be cognizant of the impact of metaphors.[309] Common metaphors may be thought of as mere figures of speech, but they "can have the effect of alienating, excluding, or seeming to disregard certain groups."[310] Burns uses the example of metaphors in which black is a negative referent, which may be awkward or offensive to African Americans.[311] Mediators should be sensitive to terms that may seem innocent but have a hurtful impact on others. An example from my own perspective is the acronym for "Jewish American Princess," "JAP." As a person of Japanese ancestry, I view that abbreviation as a homonym for a racial epithet. Stock phrases in mediation, such as "I hear what you're saying," may come across as insensitive to a hearing-impaired party. Dale Bagshaw observes that "[l]anguage is laden with social values and both carries ideas and shapes ideas."[312] Dominant discourses in Western societies tend to be Anglo-centric, as well as "agist, racist, heterosexist and homophobic."[313] Moreover, "throughout recorded history such discourses have been used by legal and social science professionals to justify categorising people as '(un)deserving,' '(ab)normal,' '(dys)functional,' '(in)competent,' '(mal)adjusted,'

---

309. Sarah E. Burns, *Thinking About Fairness & Achieving Balance in Mediation*, 35 FORDHAM URB. L.J. 39, 54 (2008). Burns' "Practice Recommendations" are associated with five general aspects of cognition: categorization (naming our world), attribution (explaining our world), metaphor (orienting our world), normative (prescribing behaviors), and framing. *Id.* at 43.

310. *Id.* at 54.

311. *Id.* (e.g., "these were dark times" or "he was one of the guys in a black hat").

312. Dale Bagshaw, *Language, Power and Mediation*, 14 AUSTRALASIAN DISP. RESOL. J. 130, 136 (2003) (citing BENJAMIN LEE WHORF, LANGUAGE, THOUGHT, AND REALITY (John B. Carroll ed., 1956)).

> Dominant dispute resolution discourses in Western cultures have tended to favour adversarial approaches to conflict and rules of law applied in formal law courts are seen as the paramount 'truths'. However, 'law' can be seen as a dominant discourse, elevated by a dominant group in a particular culture at a particular point in time, and as such can marginalise and ignore the 'truths' or ways of knowing of minority cultural groups.

Id. at 132.

313. *Id.*

'subversive,' 'delinquent' or 'deviant.'"[314] When analyzing discourses, Bagshaw notes that "[i]t is therefore crucial to identify the relationship between *what is said* and *who said it.*"[315] With this understanding, discourse analysis may reveal sexist or racist assumptions. "Language influences our attitudes and behaviour and can be used to reinforce harmful or hurtful stereotypes, such as those that are agist, sexist, racist and so forth."[316]

Bagshaw cautions mediators "to be careful in the choice of language, interpretations and the meanings they ascribe to a person's identity. Essentialism can contribute to mediators categorising and labelling clients and their problems in a way that impedes opportunities for client-centered practice and reifies and reinforces the power/knowledge of the mediator."[317] To allow parties to "supply the interpretive context for determining the meanings of events, the nature of a presenting problem, intervention and treatment,"[318] Bagshaw urges a "*reflexive* approach to [mediation] practice."[319] "In self-reflexive mediation practice it is recognised that it is impossible to be 'neutral' and the influences of characteristics such as gender, race, class, age, and sexuality on the mediator's relationship with the participants are critically examined."[320] Reflexivity demands awareness and control of the mediator's own personal and cultural biases "in order to understand the standpoint of the 'other.'"[321]

Sara Cobb and Janet Rifkin also emphasize discourse and reflexivity in their critique of mediator neutrality. They view neutrality "as a practice in discourse"[322] and assert that "existing rhetoric about neutrality does not promote reflective critical examination of discursive processes."[323] In their observations of

314. *Id.*
315. *Id.* at 136.
316. *Id.* at 137.
317. *Id.* at 139 ("Traits such as those linked to ethnicity, age, sexuality, ability or gender, should not be automatically assigned to a person's self-image as any one of those factors may not be seen by the person as relevant or important, depending on the context.").
318. *Id.*
319. *Id.*
320. *Id.* at 140.
321. *Id.* at 141.
322. Cobb & Rifkin, *supra* note 13, at 36.
323. *Id.* at 50.

Journal of Law & Policy   [Vol. 34:71

mediators, they noted that mediators "participate politically by asking questions and making summaries. Their questions bring the focus to one particular event sequence (plot), or particular story logic (theme), and/or adopt the character positions advanced by one disputant *about* another (character)."[324] In addition to mediator actions, "the structure of the mediation session itself contributes to allowing one story to set the semantic and moral grounds on which discussion and dialogue can take place."[325]

Cobb and Rifkin contend that in an effort to reduce adversarialness, mediators explore emotions, interests, fears, hopes, and needs which "obscure[] the role of discourse in the session; the mediators cannot witness their own role in the creation of alternative stories, nor can they address the colonization of one story by another."[326] The end result is that mediators contribute "to the marginalization and delegitimization of disputants."[327] For Cobb and Rifkin,

> [n]eutrality becomes a practice in *discourse*, specifically, the management of persons' positions in stories, the intervention in the associated interactional patterns between stories, and the construction of alternative stories. These processes require that mediators participate by shaping problems in ways that provide all speakers not only an opportunity to tell their story but a discursive opportunity to tell a story that does not contribute to their own delegitimization or marginalization (as is necessarily the case whenever one party disputes or contests a story in which the person is negatively positioned).[328]

Drawing on the work of Cobb and Rifkin, Isabelle Gunning describes mediation as the interaction of narratives in which the parties compete over definitions, moral positioning, and descriptions

---

324. *Id.* at 54.
325. *Id.* at 56.
326. *Id.* at 59–60. Cobb and Rifkin "recast ideology in mediation to encompass those discursive practices that privilege one story over another, that legitimize one speaker over another, that reduce any speaker's access to the storytelling process." *Id.* at 51 (citing Stuart Hall, *Signification, Representation, Ideology: Althusser and the Post-Structuralist Debates*, 2 CRITICAL STUD. IN MASS COMM. 91 (1985)).
327. *Id.* at 60.
328. *Id.* at 62.

of social relations.[329] "[T]he process of story-telling or narratives, while it has its positive aspects, may also be at the heart of the problem of bias in mediation."[330] Conversational practice is such that the first, or "primary," narrative sets the sequential and interpretive framework, and subsequent narratives are constructed in relation to that primary narrative.[331] Gunning cautions that speakers draw from the history and norms of the larger society, and when they draw on "bits and pieces of larger cultural myths" during the mediation process, "they must choose some relevant socially constructed category for themselves and others."[332] "[T]he cultural myths surrounding identity groups involving disadvantaged group members are often both negative and purely based upon derogatory conjecture and assumptions about group members."[333]

To heed these caveats about discursive practices, mediators should be extremely careful about making broad assumptions regarding a party's "culture"; "the problem with identifying 'cultural competence' as a form of neutrality is that it downplays the very real choice that mediators make in identifying 'culture.'"[334] In domestic mediations involving persons of color, generalizations about a disputant's cultural orientation based on race, ethnicity, or national origin may reflect stereotypical thinking, be over-inclusive, and be insulting to the party.[335] Cynthia Savage argues that "the common approach of defining 'culture' as being synonymous with one facet of

---

329. Isabelle R. Gunning, *Diversity Issues in Mediation: Controlling Negative Cultural Myths*, 1995 J. DISP. RESOL. 55, 68.
330. *Id.*
331. *Id.* at 68–69.
332. *Id.* at 70.
333. *Id.* at 72.
334. Clark Freshman, *Privatizing Same-Sex "Marriage" Through Alternative Dispute Resolution: Community-Enhancing Versus Community-Enabling Mediation*, 44 UCLA L. REV. 1687, 1757 (1997) (commenting on a mediation involving a Vietnamese couple and Canadian mediators where "the problematic and unspoken assumption about 'neutrality' and 'cultural competence' is that the only relevant culture is Vietnamese culture").
335. For example, a statement that "blacks might respond to the mediation context by being more expressive, using intense language as a means of communicating sincerity, or remaining fairly distant from the [white] mediator, which may increase the level of biased information coming from the disputing [black] couple" fails to account for vast differences among African Americans as individuals. *See* William A. Donohue, *Ethnicity and Mediation, in* COMMUNICATION, CULTURE, AND ORGANIZATIONAL PROCESSES 134, 147 (William B. Gudykunst et al. eds., 1985).

Journal of Law & Policy   [Vol. 34:71

cultural identity, such as race, ethnicity, or gender, is a red herring which diverts attention from the search for a more accurate and constructive approach to exploring the impact of cultural diversity on mediation."[336] Conflating culture with ethnicity may perpetuate stereotypes and ignore subcultures that contribute to an individual's cultural identity.[337]

References to culture are particularly tricky when it comes to Asian Americans who are often mistakenly thought of as natives of Asian countries instead of U.S.-born citizens. Mia Tuan's study of the "Asian ethnic experience" indicates that third- and fourth-generation Asian Americans are generally highly assimilated to white, middle-class American mainstream cultural styles and values and do not retain Chinese or Japanese cultural traditions except for commemorative events.[338] Despite this, "Asian ethnics face societal expectations to *be ethnic* since others assume they should be closer to their ethnic roots than to their American ones."[339]

Gunning exhorts us to explore cultural myths regarding disadvantaged group members in mediation through techniques such as "race-switching," or changing the races of the parties in a case study.[340] In one example, she changes the race of one character from white to Asian. In so doing, she challenges us to contend with "parts of the pre-existing narrative legitimized by the larger society, the myth that they are the 'model minority'."[341] To prevent these cultural myths from contributing to or bolstering the primary narrative, Gunning contends that mediators must "recognize that some of the cultural myths at work in the mediation process are drawn from negative taboos relating to disadvantaged groups."[342] Gunning

---

336. Cynthia A. Savage, *Culture and Mediation: A Red Herring*, 5 AM. U. J. GENDER & L. 269, 271 (1996) (proposing a "value orientation" framework as a useful way to explore cultural diversity in mediation).

337. *Id.* at 273.

338. MIA TUAN, FOREVER FOREIGNERS OR HONORARY WHITES?: THE ASIAN ETHNIC EXPERIENCE TODAY 155 (1998).

339. *Id.* at 156.

340. Gunning, *supra* note 329, at 74.

341. *Id.* at 75. Gunning observes, "Specifically, Asian-Americans of various national origins face the cultural myth of immutable foreignness . . . . There is always the question with 'foreigners' that they don't really understand 'our ways.'" *Id.*

342. *Id.* at 80.

prescribes intervention to combat negative cultural myths. Her focus on mediator intervention is a situation in which the parties interject cultural myths, urging that the mediator "may also need to flag for the parties that that is what is occurring."[343] Turning the mirror around, I urge constant vigilance and self-correction for instances when the mediator is drawing on cultural myths.

When we re-examine the Michigan mediation, we see a discursive example that disfavored and marginalized the homeowners. After the business owner presented his opening remarks and framed the dispute as a breach of contract case, the mediators questioned him in a way that reinforced his narrative. The homeowners tried to defend themselves by countering the allegation that they failed to comprehend or follow his instructions by shutting the basement door. By asking for the return of the first payment, the homeowners appeared unreasonable. If the mediators had invited the homeowners to go first in the joint session or had refrained from bolstering the carpet cleaner's narrative, the matter may have been framed as a contractor overselling his abilities and overcharging the customers.

We can imagine how the Asian American negative cultural myth of "immutable foreignness" may have bled into the Michigan mediation. In the mediators' encounter with the homeowners, the "simultaneous operation of excitatory and inhibitory cognitive processes" may have determined one category to be more dominant, and the other more suppressed.[344] If the mediators perceive the homeowners' racial category as dominant, xenophobic and race-based biases may have operated against the couple.

3. Reflexivity and Role-Playing

Bagshaw, Cobb, and Rifkin, among others, advocate reflexivity in mediation practice as a check on prejudiced subjectivity. Susan Douglas urges mediators to abandon attempts at objectivity and to instead examine one's own experiences within the mediation.[345]

---

343. *Id.*
344. C. Neil Macrae et al., *The Dissection of Selection in Person Perception: Inhibitory Processes in Social Stereotyping*, 69 J. PERSONALITY & SOC. PSYCHOL. 397, 404 (1995).
345. Douglas, *supra* note 35, at 62.

Douglas endorses reflexivity as "a useful means of conceptualising both the impact of mediator predispositions and the co-construction of meaning within the encounter."[346] Viewed this way, "reflexivity represents a rejection of mediator neutrality (in any absolute sense), an acknowledgement of the impact of mediator subjectivity and a means of addressing that subjectivity in practice."[347]

Echoing these views, Linda Mulcahy claims that her study of community mediation validates a reflexive approach in mediation.[348] Her empirical research examined one of the largest community organizations in the United Kingdom.[349] The mediators in her study admitted to having difficulty ignoring personal bias and their subjective evaluations of the merit of particular claims and parties.[350] Acknowledging these feelings, the co-mediators had debriefing sessions to discuss how their personal assessments impacted option development and process management.[351]

The Michigan mediators should have adopted reflexivity as an anti-bias method of self-assessment throughout the session. After reading the file in the small claims case, the mediators could have discussed initial reactions, assumptions, and potential issues of bias during their preparatory caucus. After the joint session, the mediators would have benefitted from a co-mediator caucus to exchange views about the parties and their respective demands. They could have made appropriate adjustments in the individual sessions to counter non-neutral thoughts and behavior. Similarly, a reflexive co-mediator discussion after each individual session may have enabled the pair to steer the mediation in a direction that was more beneficial for the parties. Even if the parties ultimately reached an impasse, they may have gained a fuller understanding of the situation and of one another's perspectives and principles. By diluting the homeowners'

---

346. *Id.* at 63 ("Reflexivity as mutual collaboration highlights the active role of the mediator in mutually reflexive dialogue . . . . Unavoidably, the mediator, rather than being a neutral facilitator of conversations, is an active coauthor in the construction of dispute narratives.").

347. *Id.* at 65.

348. Mulcahy, *supra* note 162, at 517.

349. *Id.* at 515.

350. *Id.* at 516.

351. *Id.* at 517.

narrative and failing to explore a range of options, the mediators legitimized the business owner's version of entitlement.

Through the use of various practices, we attempt to incorporate discursive and reflective theories in a law school mediation clinic. As Cobb, Rifkin, and Gunning make apparent, the party who speaks first has the advantage of painting a subjective picture of the circumstances underlying the dispute.[352] Student-mediators are eager to ask a litany of "fact-gathering" questions (to the point of interrogation) to the first speaker before listening to the second speaker's narrative. By doing so, students add their own "spin" and make assumptions that may be tainted by their own experiences and expectations. Thus, they may re-characterize or validate the first speaker's presentation through their own additions. Rather than presenting her own "story," the second speaker is reduced to opposing a pre-determined version embellished by the mediators. This can frustrate and incite defensiveness in the second speaker who has been asked to wait her turn and not interrupt.

Recognizing this dynamic, students are directed to refrain from asking questions until both parties have had the opportunity to supply their narratives in their own words and styles. In what may be an atypical practice, we refrain from summarizing and reframing the first person's statements before the second person speaks. While there is always some perceived favoritism that one party goes first, withholding questions and postponing summarizing or reframing lessens the likelihood that the second speaker's narrative will be molded by others.

It is also important to model lack of bias in selecting the party who speaks first. Asking the parties who would like to go first may be perceived as rewarding one party over the other (the more assertive party or the one closest to the mediator, for example). Mediators evidence external neutrality by being transparent in decision-making. Parties should be told why and how the mediators determined the order of presentations (for example, a random method of selection, such as by alphabetical order or coin toss).

---

352. Gunning, *supra* note 329, at 68–70.

Reflective learning has been described as "an intentional social process, where context and experience are acknowledged, in which learners are active individuals, wholly present, engaging with others, open to challenge, and the outcome involves transformation as well as improvement for both individuals and their environment."[353] Like many clinical legal educators, I have included role-playing exercises as a reflective teaching opportunity in the mediation clinic for decades.[354] Role-plays contain the essential elements of learning and reflection: (1) "a genuine situation of experience"; (2) a "genuine problem in that situation"; (3) "information and observation about the situation"; (4) "suggested solutions for which the student [is] responsible"; and (5) "opportunity . . . to test ideas by application."[355] By practicing in an academic setting, students will (hopefully) transfer the lessons to their actual cases.

My mediation clinic students participate in five increasingly difficult two-hour role-plays as parties, co-mediators, and observers.[356] In addition to helping the students to improve their mediation skills, the role-plays enable the students to develop empathy and view the process from the perspective of the disputants. Students are encouraged to experiment and put ideas into action. During the role-plays, mediators explore their decision-making processes, assess progress, and consider their reactions to options. Mediators are asked to express how their thoughts and feelings motivated them and evaluate to what extent they pushed options.[357] During class discussion, we deconstruct the mediation role-play and

353. Samantha Hardy, *Teaching Mediation as Reflective Practice*, 25 NEGOTIATION J. 385, 389 (2009) (quoting ANNE BROCKBANK & IAN MCGILL, FACILITATING REFLECTIVE LEARNING IN HIGHER EDUCATION 36 (2d ed. 2007)).

354. For a critique of role-plays as a learning activity, see Nadja Alexander & Michelle LaBaron, *Death of the Role-Play*, *in* RETHINKING NEGOTIATION TEACHING: INNOVATIONS FOR CONTEXT AND CULTURE 179, 179–97 (Christopher Honeyman et al. eds., 2009).

355. Hardy, *supra* note 353, at 390 (citing BROCKBANK & MCGILL, *supra* note 353, at 23).

356. By this time, they have also viewed a small claims mediation at the courthouse, observed an in-class mock mediation demonstration by experienced mediators, and engaged in skills development exercises.

357. Hardy, *supra* note 353, at 397 (quoting MICHAEL D. LANG & ALISON TAYLOR, THE MAKING OF A MEDIATOR: DEVELOPING ARTISTRY IN PRACTICE 54 (2000)) ("Elicitive questioning" presses "mediators to uncover for themselves what was successful or unsuccessful, and to identify the reasoning behind their strategies and approaches, and . . . consider the impact of their interventions on the disputants.").

all students offer oral comments. Both the students and the instructors also complete written critiques.[358] The purpose of feedback is not merely to give mediators a "360 degree" evaluation but also to allow the students and instructors to engage in collective problem-solving and to examine assumptions and reactions.[359] We record the role-plays and make them available for students to view on their laptops. By watching their performances and the reactions of the parties, student mediators can see or contest the validity of the feedback that was offered. They can also see if they were guilty of behavior that reflected bias or favoritism, such as facial expressions, body language, or blinking.

Mediation teachers and trainers should answer Gunning's call for more deliberate confrontation of racial stereotypes and assumptions in training.[360] With that goal in mind, I designed a role-play simulation based on community tensions in Washington, D.C., for use in my mediation course.[361] It is a composite of disputes arising out of years of ongoing tension between Korean American shopkeepers and African American customers.[362] I have varied my approach over time. I initially played the shopkeeper and later opted to recruit volunteers from student groups to play the disputants. For

---

358. Students complete evaluations as mediators, parties, and observers.

359. Hardy, *supra* note 353, at 393 (quoting BROCKBANK & MCGILL, *supra* note 353, at 5). In this way, "learners and teacher engage and work together so that they jointly construct meaning and knowledge from the material." *Id.*

360. Gunning, *supra* note 329, at 86–88.

361. In fact, as a participant at a conference hosted by the UCLA Center for Study and Resolution of Interracial/Interethnic Conflict, March 28–30, 1996, Professor Gunning offered constructive comments to refine the role-play. For an analysis of Black-Korean tension, see Kyeyoung Park, *Use and Abuse of Race and Culture: Black-Korean Tension in America, in* THE CONFLICT AND CULTURE READER 152, 152–62 (Pat K. Chew ed., 2001).

362. *See, e.g.,* Michael A. Fletcher, *Asian-Owned Carryout is Focus of Rally: Small Group Protests Nonblack Business,* WASH. POST, Oct. 19, 1996, at D4 (reporting on African American protestors engaged in a protest rally outside of an Asian-American restaurant in D.C.); *see also Mayor's Proposed Fiscal Year 2010 Budget: Before the Committee on Aging and Community Affairs,* Apr. 24, 2009 (statement of Francey Youngberg, Chair, DC Fair Access Coalition) ("According to the *Washington Post,* two-thirds of all business licenses are owned by Asian Pacific Americans in the District. D.C. agencies estimate that 60% of corner groceries and 57% of lotteries are sold through Asian-owned stores.") (copy on file with author). According to the 2000 U.S. Census, roughly 60 percent of D.C. residents are black. *District of Columbia-DP-1. Profile of General Demographic Characteristics: 2000,* CENSUS.GOV, http://factfinder.census. gov/servlet/QTTable?_bm=y&-geo_id=04000US11&-qr_name=DEC_2000_SF1_U_DP1&ds_ name (last visited Nov. 8, 2010).

the past few years, I have shown a videotape demonstration; this technique has the advantage of making the discussion about process dynamics and co-mediator choices easier. The demonstration allows students to discuss stereotypes, prejudice, and bias in a controlled and confidential setting. We also use other role-plays and scenarios that involve racial or gender dynamics.

4. Co-Mediation and Race-Matching

Co-mediation offers a number of advantages for advancing external neutrality.[363] We use a co-mediation model exclusively in our community program.[364] Having "two heads" allows the co-mediator team to engage in an explicit discussion of how "neutrally" they are operating within a particular mediation context.[365] In co-mediator caucuses, the team can engage in active reflection to assess the discursive dialogues, interactions with and between disputants, and inclinations to favor or disfavor options. Rather than rushing to an agreement on approach and actions, co-mediators can play "devil's advocate" to affirmatively critique their behavior and choices. A co-mediator provides the eyes and ears for peer evaluation. Although mediators may be reluctant to offer constructive criticism (since it is not anonymous), a mutual co-mediator evaluation can incorporate elements of debriefing, reflection, positive feedback, and suggestions for future improvement. These co-mediator assessments would provide a useful supplement to party evaluations, which are employed by most mediation programs. Peer evaluation of mediators could be accomplished in other ways. For example, the D.C. Superior Court Multi-Door Dispute Resolution Branch uses a one-way mirror so evaluators can observe mediations without being seen by the participants.

---

363. Gunning, *supra* note 329, at 88–89 (citing the benefits of using mediator teams to combat negative cultural myths).

364. Students in my Consumer Mediation Clinic are sole mediators of consumer-business disputes, whereas students in my Community Dispute Resolution Center Project co-mediate adult misdemeanor, juvenile delinquency, and police-civilian disputes.

365. For good suggestions on making the most of co-mediation, see Lela P. Love & Joseph B. Stulberg, *Practice Guidelines for Co-Mediation: Making Certain That "Two Heads Are Better Than One,"* 13 MEDIATION Q. 179 (1996).

Co-mediation also leverages differences in perspectives and experiences when you have mediators of different ethnicities, genders, or abilities.[366] This can provide a check on biased and discriminatory mediator actions. For example, a female mediator might help her male partner avoid gendered comments and assumptions. In some mediation contexts, pairing mediators is done deliberately and strategically to create complementary duos.[367] Advocates of "race-matching" co-mediator teams to mirror the racial or ethnic distribution of the parties cite several benefits: symbolic fairness, increased likelihood that mediators and parties will have shared experiences, modeling equality, and broader interpretive frameworks.[368]

Clark Freshman points out several dangers of matching parties with mediators based on common traits or affiliations.[369] "First, psychologists have found it notoriously difficult to predict precisely how individuals, be they mediators or not, will see some as 'we' and others as 'they.'"[370] Second, there may be biases within individual communities. "Leading psychologists of discrimination suggest that, as much as we think we know how others see themselves, individuals may divide the world in many different ways."[371] He adds that "[a] reciprocal problem may arise when some who identify strongly with a community have negative views of those who they feel have betrayed their 'true' identity by trying to assimilate or fit some other community instead."[372] Another problem with matching mediators is that the practice may exacerbate discrimination outside the community.[373] This operates in two ways: positive contact with

---

366. Gunning, *supra* note 329, at 88–89.
367. For example, in emotional family disputes, a team containing a lawyer and therapeutic counselor might be beneficial. In a heterosexual divorce, a male and female mediator team might be used. *Id.* at 88.
368. *Id.* at 89.
369. Clark Freshman, *The Promise and Perils of "Our" Justice: Psychological, Critical and Economic Perspectives on Communities and Prejudices in Mediation*, 6 CARDOZO J. CONFLICT RESOL. 1, 10–11 (2004).
370. *Id.* at 10.
371. *Id.* at 11. Freshman uses the example of relatively assimilated Jews who "may often express more negative views about those not assimilated than even the most inside group." *Id.* at 12.
372. *Id.*
373. *Id.*

dissimilar persons often reduces prejudice, and unconscious bias against the group may become less prevalent.[374]

Furthermore, race-matching risks essentialism and reflects reductionist assumptions about individuals.[375] Amartya Sen describes a kind of reductionism that he refers to as "singular affiliation."[376] This reductionism "takes the form of assuming that any person preeminently belongs, for all practical purposes, to one collectivity only—no more and no less. Of course, we do know in fact that any real human being belongs to many groups, through birth, associations, and alliances."[377] Individuals should be able to choose which affiliations are more relevant or important in any social context and not have others impose that on them; political affiliation or religion, for example, may trump race.[378] Race-matching the mediators for parties of Asian descent ignores ethnic, national, regional, political, religious, socio-economic, and other differences that may be more relevant or important in a given situation than shared racial category.[379]

Finally, a study of race-matching revealed that "[w]ith regard to mediation outcomes . . . it is not so clear that creating racial matches between mediation participants and mediators is as important as we have thought in the past."[380] A multiyear research project in Maryland community mediation centers determined that

> when the mediator is not of the same race as either participant, participants believe that they have been heard by the mediator. In contrast, when the mediator's race matches that of the opposing party, the participant is less likely to feel that the

---

374. *Id.* at 12–13. Freshman also notes that matching could "trigger the unconscious stereotype that 'they' are clannish." *Id.* at 13. Moreover, "even if one adopts the less separatist notion of teaching cultural 'sensitivity' to mediators . . . the 'sensitivity' may harden the way mediators automatically divide the world into group terms." *Id.*

375. *See* Bagshaw, *supra* note 312, at 139 (discussing the dangers of essentialism and assigning possibly irrelevant traits to a person's self-image).

376. AMARTYA SEN, IDENTITY AND VIOLENCE: THE ILLUSION OF DESTINY 20 (2006).

377. *Id.*

378. *Id.* at 29–32.

379. Bagshaw, *supra* note 312, at 139.

380. Lorig Charkoudian & Ellen Kabcenell Wayne, *Does It Matter If My Mediator Looks Like Me? The Impact of Racially Matching Participants and Mediators*, DISP. RESOL. MAG., Spring 2009, at 22, 24.

mediator listened to her. A similar negative effect occurs with regard to participants' sense of control over the conflict situation. This sense of control does not change when the mediators' race is different from both participants, but decreases from the beginning to the end of the mediation when the mediator's race matches only that of the opposing party. Again, it appears less important to have a mediator who 'looks like me' than it is to avoid having a mediator who 'looks like' the other participant and no mediator who 'looks like me.'[381]

The researchers suggest that their finding supports "the value of co-mediation, which creates more options for addressing racial balance amongst participants and mediators."[382]

5. Transformative Mediation and Procedural Justice

Transformative mediation and procedural justice theories suggest that external neutrality would be improved through a process that ensures a high degree of control for the disputants. Joseph Folger and Robert Baruch Bush postulate that neutrality is unachievable because the mediator's interests become part of the problem-solving endeavor; they propose that their transformative model ensures party self-determination.[383] They contend that a problem-solving mediation, which focuses on reaching agreement, "leads mediators to be directive in shaping both the problems and the solutions, and they wind up influencing the outcome of mediations in favor of settlement generally and in favor of terms of settlement that comport with their views of fairness, optimality, and so forth."[384] In transformative mediation, "[n]eutrality means that the mediator's only interest is the interest in using his or her influence to make sure that *the parties maintain control* of decisions about outcomes."[385]

---

381. *Id.* at 24.
382. *Id.*
383. ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: THE TRANSFORMATIVE APPROACH TO CONFLICT 22–26, 105 (rev. ed. 2005).
384. *Id.* at 104.
385. *Id.* at 105. Astor also endorses an approach that emphasizes self-determination, party empowerment, and collaboration between the parties. Astor, *supra* note 11, at 78.

Mary Beth Howe and Robert Fiala analyzed randomly assigned small claims court mediation cases in New Mexico to evaluate factors affecting disputant satisfaction with mediation.[386] Data from the study show that certain factors in the mediator's control are strongly associated with party satisfaction. For example, party satisfaction increases when "the mediator appears neutral, is in control of the mediation, and allows participants to feel they are able to tell their story. Greater participant integration, less anger and hostility, and greater power in mediation are also linked to satisfaction."[387] Structural factors associated with social class, gender, and ethnicity showed "few and inconsistent links to satisfaction."[388]

In her exegesis of procedural justice literature, Rebecca Hollander-Blumoff identifies four dominant factors in assessments of process fairness: "opportunity for voice, courteous and respectful treatment, trustworthiness of the decision-maker, and neutrality of the decision-maker."[389] Procedural justice legitimizes the mediation process and increases the likelihood that the outcome will be accepted by the participants. In their well-known compilation of studies of dispute resolution systems, John Thibaut and Laurens Walker concluded that "the maintenance of a high degree of control . . . by disputants and, at the same time, . . . a high degree of regulated contentiousness between the disputants themselves" are important properties for a just procedure.[390] Their research revealed "that a procedure that limits third-party control, thus allocating the preponderance of control to the disputants, constitutes a just procedure."[391]

In short, we can draw upon multiple lessons to check external neutrality. Neutrality is promoted by managing the mediation process to maintain even-handed, respectful treatment of disputants and by

---

386. Mary Beth Howe & Robert Fiala, *Process Matters: Disputant Satisfaction in Mediated Civil Cases*, 29 JUST. SYS. J. 85 (2008).

387. *Id.* at 93.

388. *Id.* at 94.

389. Rebecca Hollander-Blumoff, The Theoretical and Empirical Case for Procedural Justice in Negotiation 9 (Sept. 9, 2009) (unpublished manuscript) (on file with author).

390. JOHN THIBAUT & LAURENS WALKER, PROCEDURAL JUSTICE: A PSYCHOLOGICAL ANALYSIS 119 (1975).

391. *Id.* at 118.

maximizing party control. In addition, mediators can attend to external neutrality concerns by: being sensitive to language usage; valuing individual party narratives; ensuring that disputants "tell their stories" in their own words and style; self-policing for essentialist assumptions; and monitoring for biased party interventions. Finally, adopting a reflexive approach that is deliberatively self-conscious; using co-mediator teams that leverage differences and similarities; and employing instructional methods that require mediators to grapple with racial and other difficult issues would further reduce the potential for mediator partiality and bias.

B. Internal Neutrality

Having identified steps a mediator may undertake to address external neutrality issues, we now look inward to consider what mediators can do to minimize the operation of biased mental processes that are automatic and not a part of our conscious awareness. Research shows that suppression of stereotyped associations and engagement of non-prejudiced responses requires "intention, attention, and effort."[392] Fortunately, mediators have the power and ability to improve internal neutrality measures to reduce bias and favoritism in mediation. Practical suggestions include setting goals, planning deliberate actions to reduce biased responses, increasing diversity of mediator contacts, applying mindfulness techniques, and developing a habit of practices that remove bias.

1. Awareness, Motivation, and Action

Awareness of bias is critical for mental decontamination success.[393] As one may expect, the first step toward internal neutrality is to acknowledge the existence of unconscious mediator biases and

---

392. Armour, *supra* note 144, at 24 (quoting Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 16 (1989)).

393. Laurie A. Rudman et al., *"Unlearning" Automatic Biases: The Malleability of Implicit Prejudice and Stereotypes*, 81 J. PERSONALITY & SOC. PSYCHOL. 856, 866 (2001) (citing Timothy D. Wilson & Nancy Brekke, *Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations*, PSYCHOL. BULL. 117, 117–42 (1994)).

Journal of Law & Policy   [Vol. 34:71

prejudices. "[I]n order to counter otherwise automatic behavior, one must accept the existence of the problem in the first place. . . . We must be both aware of the bias and motivated to counter it. If we instead trust our own explicit self-reports about bias—namely, that we have none—we will have no motivation to self-correct."[394] Requiring mediators to take the IAT for an implicit bias "reality check" could potentially open their eyes to their own egalitarian shortcomings. I ask my students to do such an exercise early in the semester.[395] Although it is voluntary, I request that they take one of the implicit association tests and complete a questionnaire anonymously.[396] During the semester, we reflect on and refer back to the experience as it relates to actual cases.

When confronted with their own implicit attitudes and stereotypes, mediators can work to counter the operation of bias. With increased awareness of implicit bias and the goals and motivation to self-correct, mediators can begin to tackle the problem of unintentional unequal treatment of parties. Researchers found that merely knowing one's prejudice level was not sufficient to respond in a less prejudiced manner.[397] People who are externally motivated (wanting to appear non-prejudiced to other people) to reduce prejudice-related reactions are more likely to adjust a prejudiced act based on the social context they are in, while those who are only internally motivated (appearing non-prejudiced to oneself) may not be so affected by social pressures.[398] It is possible "that external motivation precedes internal motivation and that to initiate change, the social climate must discourage expressions of prejudice."[399] The

---

394. Kang, *supra* note 136, at 1529.

395. I got this idea from Gary Blasi, who posted an e-mail on the clinical list serve on August 1, 2007, in response to Gail Silverstein's inquiry about incorporating the IAT in clinic courses. Blasi explained that he has used the IAT, but he always used it in conjunction with reading and discussion of the science behind the IAT and the implication for lawyers." E-Mail from Gary Blasi, Professor of Law, UCLA Sch. of Law, to Gail Silverstein, Clinical Att'y, Civil Justice Ctr., Univ. of Cal. Hastings Coll. of Law (Aug. 1, 2007, 12:21:53 PST) (on file with author).

396. The simple questionnaire asks for their reactions and reflections on the test experience and their "scores."

397. E. Ashby Plant & Patricia G. Devine, *Internal and External Motivation to Respond Without Prejudice*, 75 J. PERSONALITY & SOC. PSYCHOL. 811, 826 (1998).

398. *Id.* at 825.

399. *Id.* at 827.

researchers observed that "although discouraging overtly prejudiced responses may be desirable, it appears that internal motivation may be necessary to sustain efforts to respond without prejudice over time, particularly when no immediate external standards are salient."[400]

Recent studies show that while stereotypes may be automatically *activated*, as conscious actors we may be able to affect the *application* of those stereotypes in our interactions, judgments, and decisions. Irene Blair and Mahzarin Banaji conducted a series of four experiments to observe the automatic activation of gender stereotypes and to assess conditions under which stereotype priming may be moderated.[401] They distinguish between stereotype activation (categorization) and stereotype application as sequential steps in the process. They believe that stereotype activation is an automatic process, whereas stereotype application is a controlled, or at least a controllable, process.[402] Their experiments revealed that even with the "strong and ubiquitous nature of stereotype priming, . . . such effects may be moderated under particular conditions. . . . [s]tereotype priming can be eliminated when perceivers have an intention to process counterstereotypic information and sufficient cognitive resources are available."[403]

In another experiment on reducing the application of stereotypes, Margo Monteith observed that low prejudiced individuals experienced prejudice-related discrepancies (i.e., a prejudiced response such as feeling uncomfortable sitting next to a gay male on a bus) even though they believed the response was inappropriate.[404] She investigated whether people can inhibit prejudiced responses and

400. *Id.* (citing David P. Ausubel, *Relationships Between Shame and Guilt in the Socializing Process*, 62 PSYCHOL. REV. 378, 378–90 (1955)). Later studies determined the importance of internal motivation, finding that the measure of implicit bias was lowest among individuals with high levels of internal motivation and low level of external motivation. *See* Patricia G. Devine et al., *The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice*, 82 J. PERSONALITY & SOC. PSYCHOL. 835 (2002).

401. Irene V. Blair & Mahzarin R. Banaji, *Automatic and Controlled Processes in Stereotype Priming*, 70 J. PERSONALITY & SOC. PSYCHOL. 1142 (1996).

402. *Id.* at 1143.

403. *Id.* at 1159.

404. Margo J. Monteith, *Self-Regulation of Prejudiced Responses: Implications for Progress in Prejudice-Reduction Efforts*, 65 J. PERSONALITY & SOC. PSYCHOL. 469, 469 (1993).

respond on the basis of personal non-prejudiced beliefs.[405] She found that the discrepancy experience produced a negative self-directed effect which increased motivation for discrepancy reduction.[406] Increased attention to discrepancy-relevant information and personal discrepancy experiences may help low prejudiced individuals exert control over their biased responses.[407] Importantly, she found that "prejudice-related discrepancy experience enabled the low prejudiced subjects to be more effective at inhibiting prejudiced responses at a later time."[408]

In addition to recognizing implicit bias and having adequate motivation to reduce it, mediators must call upon cognitive control processes. Blair and Banaji's experiments examined the automatic processes underlying stereotyping and the role of intention and cognitive resources in moderating the influence of such processes on one's judgment.[409] The results suggest that people can control or eliminate the effect of stereotypes on their judgments if they have the intention to do so and their cognitive resources are not over-constrained.[410] After reviewing numerous studies, Blair discovered that automatic stereotypes are influenced by social and self-motives, specific strategies, the perceiver's focus of attention, and the configuration of stimulus cues.[411] In a study by Bruce Bartholow and colleagues, participants drinking alcohol showed significantly impaired regulative cognitive control and diminished ability to inhibit race-biased responses, suggesting that controlling racial bias can be a function of implementing cognitive control processes.[412]

With the requisite motivation and cognitive resources to draw upon, mediators are ready to operationalize a bias reduction plan. Gollwitzer, Sayer, and McCulloch propose "implementation-

---

405. Id. at 472.
406. Id. at 477.
407. Id. (citing JEFFREY A. GRAY, THE NEUROPSYCHOLOGY OF ANXIETY: AN ENQUIRY INTO THE FUNCTIONS OF THE SEPTOHIPPOCAMPAL SYSTEM (1982)).
408. Id. at 482.
409. Blair & Banaji, supra note 401, at 1142.
410. Id. at 1159.
411. Blair, supra note 85, at 242.
412. Bruce D. Bartholow et al., Stereotype Activation and Control of Race Bias: Cognitive Control of Inhibition and Its Impairment by Alcohol, 90 J. PERSONALITY & SOC. PSYCHOL. 272 (2006).

intention" as an approach to situations that may trigger implicit bias responses.[413] Goal-intention is expressed as "I intend to reach X goal."[414] Goal-directed behavior is important, but might not become part of everyday routine. "As a substitute, people can resort to forming implementation-intentions that strategically place the intended goal-directed behavior under direct situational control."[415] Implementation-intention is expressed as "if X, then I will do Y."[416] Implementation intentions are expressed as plans to reach the goal.[417] Implementation intention studies have shown promising results: participants were generally more likely to attain their goals, were more resistant to distracters, and showed less stereotype activation.[418]

Applying these strategies to mediation clinics and programs, instructors and administrators should articulate explicit program goals and guidelines about expected mediator non-prejudiced behavior and incentivize actions to meet those goals. In one example of an interesting innovation, the American Bar Association (ABA) has offered a continuing legal education program on "Creating a Culture of Inclusion" and made available "Elimination of Bias Credit."[419] In lieu of the typical pro forma "diversity" segment in mediation trainings, teachers and trainers should consider a more robust anti-prejudice curriculum. Gunning advocates inclusion of "misperceptions of different identity groups as part of the mediation training. These discussions and explorations would and should be a

---

413. Peter M. Gollwitzer et al., *The Control of the Unwanted, in* THE NEW UNCONSCIOUS 485, 486–87 (Ran R. Hassin et al. eds., 2005).
414. *Id.* at 487.
415. *Id.* at 486.
416. *Id.* at 486–87.
417. Gollwitzer and his colleagues use this example:

When participants had furnished their goal intentions of judging the elderly in a nonstereotypical manner with the respective implementation intention ("If I see an old person, then I will tell myself: Don't stereotype!"), the typical automatic activation of stereotypical beliefs . . . was even reversed. Similarly, when participants had the goal intention to judge female job applicants in a nonstereotypical manner and furnished an implementation intention to ignore a certain applicant's gender, no automatic activation of stereotypical beliefs about the female was observed.

Id. at 495.
418. *Id.* at 496.
419. Am. Bar Ass'n Ctr. for Continuing Legal Educ., *Creating a Culture of Inclusion*, AM. BAR ASS'N, http://www.abanet.org/cle/programs/t10cci1.html (last visited Nov. 8, 2010).

part of the basic mediation training not relegated as they so often are to some advanced form of training on 'cross-cultural mediation' or 'how to deal with power-imbalances'."[420] Mandatory continuing mediation education for mediators practicing in particular programs or jurisdictions could include "elimination of bias" credits and certification of anti-bias coursework.

In addition to the normative (external) incentive, mediators must set their own personal (internal) goals of egalitarianism. A general aspiration to be "neutral" is insufficiently specific. To achieve more fairness in mediation, Burns recommends that mediators affirm that their goal is to be fair and non-discriminatory.[421] She also urges mediators to monitor how they make distinctions and to assume they are biased in favor of members of their own group and against persons in other groups.[422] Internally motivated mediators should develop their own "intention-implementation plans" for goal attainment and tailor them for specific mediation settings. As part of pre-mediation preparation, mediators should consider potential bias pitfalls that might arise in interracial disputes and develop reaction plans to avoid or escape the traps.

2. Salience, Exposure, and Practice

Racially discriminatory behavior may be reduced more effectively when racial issues are made salient rather than ignored or obscured.[423] Research shows that focusing attention on the source of a possible implicit effect that interferes with judgment reduces or eliminates (or even reverses) the interference.[424] For example, the false fame effect was reduced when sufficient attention was focused on the initial list of non-famous names so the subjects would recognize non-famous names as having been encountered earlier in

---

420. Gunning, *supra* note 329, at 87.
421. Burns, *supra* note 309, at 44.
422. *Id.* at 45. "[E]ven if one somehow has been consciously oblivious to the presence of key social differences, failing to consider the effects of social difference is the strategy most likely to perpetuate historic patterns of bias." *Id.* at 50.
423. Wang, *supra* note 210, at 1038 (citing Samuel R. Sommers & Phoebe C. Ellsworth, *White Juror Bias: An Investigation of Prejudice Against Black Defendants in the American Courtroom*, 7 PSYCHOL. PUB. POL'Y & L. 201, 220–21 (2001)).
424. Greenwald & Banaji, *supra* note 67, at 18.

the experiment.[425] "Drawing social category information into conscious awareness allows mental (cognitive and motivational) resources to overrule the consciously unwanted but unconsciously operative response."[426]

Taking a similar view, Jody Armour agrees that decision-makers would be more likely to become aware of their implicit biases, confront them, and hopefully counteract their effects when such references are explicitly made.[427] Citing the distinction between a habit (an automatic process done many times) and a decision (a conscious action), Armour proposes that "for a person who rejects the stereotype to avoid stereotype-congruent [behavior] responses to blacks (i.e, to avoid falling into a bad habit), she must intentionally inhibit the automatically activated stereotype and activate her newer personal belief structure."[428] Since people may act on stereotypes automatically and without knowledge, they must actively monitor and inhibit the automatic stereotype and replace it with a personal egalitarian belief.[429] "[U]nless a low-prejudiced person consciously

---

425. *Id.* (citing Larry L. Jacoby et. al., *Becoming Famous Overnight: Limits on the Ability to Avoid Unconscious Influences of the Past*, 56 J. PERSONALITY & SOC. PSYCHOL. 326 (1989)).

426. Banaji & Greenwald, *supra* note 68, at 70. Some studies, however, imply that stereotypes are more difficult to suppress through controlled processes. In an experiment that required subjects to make a judgment of criminality using names that vary racially (black, white, Asian), researchers found race bias was difficult to remove even when subjects were alerted that racist individuals are more likely to identify black compared to white names. *See* Banaji & Dasgupta, *supra* note 133, at 162. Another study showed that participants "explicitly instructed to *avoid using race* ironically performed worse (although not in a statistically significant way) than participants told nothing at all." Kang, *supra* note 136, at 1529 (citing B. Keith Payne et al., *Best Laid Plans: Effects of Goals on Accessibility Bias and Cognitive Control in Race-Based Misperceptions of Weapons*, 38 J. EXPERIMENTAL SOC. PSYCHOL. 384, 390–91 (2002)). Researchers have also observed an effect called "stereotype rebounding." When people attempt to repress stereotypic thoughts, these thoughts may subsequently reappear with even greater insistence and be even more difficult to ignore. C. Neil Macrae et al., *Out of Mind but Back in Sight: Stereotypes on the Rebound*, 67 J. PERSONALITY & SOC. PSYCHOL. 808 (1994).

Thought suppression operates by searching for a distracter to replace the unwanted thought; however, when cognitive resources are limited, the ability to search for a distracter is precluded and the unwanted thought becomes hyperaccessible. *Id.* at 809 (citing Daniel M. Wegner & Ralph Erber, *The Hyperaccessibility of Suppressed Thoughts*, 63 J. PERSONALITY & SOC. PSYCHOL. 903, 903–12 (1992)).

427. Armour, *supra* note 144, at 13.

428. *Id.* at 24.

429. *Id.* at 23–24.

monitors and inhibits the activation of a stereotype in the presence of a member (or symbolic equivalent) of a stereotyped group, she may unintentionally fall into the discrimination habit."[430] Importing this model to the mediation context, mediators must break the habit of stereotype-consistent behavior by making conscious decisions to act in accordance with their non-discriminatory beliefs.

We use mediation debriefings, case rounds, and journals to give students in the mediation clinic space within which to contemplate and comment on their reactions to situations in which prejudiced behavior and assumptions could, or did, surface. More importantly, students identify lessons they can take into future mediations. The practice of journaling, which is popular in law school clinics, is a learning device that would benefit veterans as well as new mediators. This type of written reflection could be adapted to court and community settings to encourage mediators to measure adherence to their own egalitarian goals throughout their mediations. Requiring mediators to articulate explicit plans for improvement challenges them to name their practice shortcomings and state personal performance goals and intentions. Administrators of mediation programs should embrace these activities by periodically bringing volunteers and staff together for candid conversations and brainstorming sessions on prejudice reduction strategies. Inexpensive "brown bag" lunch discussions on a regular basis would be a cost- and time-effective way to help mediators take basic steps toward bias reduction.

Implicit social cognition research indicates that bias can be reduced through exposure to individuals who are not like us.[431] This exposure can occur through interpersonal interaction and presentation of images. The "Social Contact Hypothesis" postulates that stereotypes and prejudice can be reduced when people of different social categories have face-to-face interaction under certain conditions.[432] A recent meta-analysis of studies found that intergroup contact correlates negatively with prejudice.[433] Intergroup contact

---

430. *Id.* at 24.
431. Kang & Banaji, *supra* note 140, at 1101.
432. *Id.*
433. *Id.* at 1102–03.

may actually reduce levels of implicit bias.[434] In one study, white subjects were asked to "take the race IAT and report the number of their close out-group friends: African-Americans in one experiment and Latinos in another. . . . The researchers found negative correlations between the number of interracial friendships and level of implicit bias."[435] Consistent with these findings, the C100 study referenced earlier revealed that "the more prejudiced respondents tend to interact less frequently with Chinese and Asian Americans."[436]

Along similar lines, implicit attitudes may be changed by exposure to positive images.[437] In one study, subjects were shown photos of Martin Luther King Jr. and Denzel Washington as positive Black images.[438] The group reduced implicit bias by more than half and the effect persisted for a full day.[439] In the same manner, counter-typical visualizations caused a decrease in implicit stereotypes in another experiment.[440] In an experiment on explicit and implicit bias against women, direct educational instruction by counter-typical exemplars (female faculty) over one year had significant decreasing effects on IAT scores.[441]

These research findings suggest that mediators may be able to reduce implicit bias through increased exposure to and encounters with positive examples of out-group members. Writing about racial

---

434. Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 CALIF. L. REV. 969, 981 (2006) ("A significant body of social science evidence supports the conclusion that the presence of population diversity in an environment tends to reduce the level of implicit bias.").

435. Kang & Banaji, *supra* note 140, at 1103 (citing Christopher L. Aberson et al., *Implicit Bias and Contact: The Role of Interethnic Friendships*, 144 J. SOC. PSYCHOL. 335, 340, 343 (2004)).

436. COMMITTEE OF 100 & HARRIS INTERACTIVE, *supra* note 262, at 68.

437. Nilanjana Dasgupta & Anthony G. Greenwald, *On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals*, 81 J. PERSONALITY & SOC. PSYCHOL. 800 (2001).

438. *Id.* at 802.

439. *Id.* at 807; *see also* Irene V. Blair et al., *Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery*, 81 J. PERSONALITY & SOC. PSYCHOL. 828, 837 (2001); Nilanjana Dasgupta & Shaki Asgari, *Seeing Is Believing: Exposure to Counterstereotypic Women Leaders and Its Effect on the Malleability of Automatic Gender Stereotyping*, 40 J. EXPERIMENTAL SOC. PSYCHOL. 642 (2004).

440. Kang & Banaji, *supra* note 140, at 1107 (citing Blair et al., *supra* note 439, at 828–29).

441. *Id.* (citing Dasgupta & Asgari, *supra* note 439, at 651).

issues in mediation, Howard Gadlin decries the lack of diversity in the dispute resolution field and urges greater racial and ethnic integration.[442] Homogeneity among mediator ranks has spurred efforts to increase the numbers of minorities and expand practice opportunities for mediators of color.[443] Employing counter-typical mediation trainers and teachers and enlarging mediator diversity would be rational moves toward implicit bias reduction. Because in-group favoritism makes it hard to reduce prejudice, Carwina Weng notes that mere interaction with other groups is insufficient. Contact in a setting that promotes equality and openness is critical.[444] She lists cooperation, constructive conflict resolution and internalized civic values as elements for building an egalitarian community in which non-discriminatory relationships are fostered.[445]

A diversity training experiment supports Weng's suggestion that prejudice reduction is more successful when interaction is coupled with supporting knowledge and efforts. Researchers found that students enrolled in a prejudice and conflict seminar taught by an African American male professor were able to lower their bias by the end of the semester.[446] Specific data indicated that an "[i]ncreased awareness of discrimination against African Americans and motives to overcome prejudice in oneself" was more correlated with a reduction in explicit bias, while a "positive evaluation of the professor and the prejudice and conflict seminar," making friends with out-group members, and reporting feeling less threatened by out-group members, were more correlated with a reduction in implicit prejudice and stereotyping.[447] A control group taught by an African American professor showed no reduction in prejudice and bias, leading to the conclusion that the presence of an African American

---

442. Howard Gadlin, *Conflict Resolution, Cultural Differences, and the Culture of Racism*, 10 NEGOTIATION J. 33, 44 (1994).

443. Marvin E. Johnson & Homer C. La Rue, *The Gated Community: Risk Aversion, Race, and the Lack of Diversity in Mediation in the Top Ranks*, DISP. RESOL. MAG., Spring 2009, at 17.

444. Carwina Weng, *Individual and Intergroup Processes to Address Racial Discrimination in Lawyering Relationships*, in CRITICAL RACE REALISM: INTERSECTIONS OF PSYCHOLOGY, RACE, AND LAW 64, 70 (Gregory S. Parks et al. eds., 2008).

445. *Id.* at 73.

446. Rudman et al., *supra* note 393, at 856.

447. *Id.* at 865 tbl.7.

figure in a prominent position alone had little to no effect on implicit or explicit bias.[448]

3. Mindfulness Meditation

A growing number of dispute resolution scholars tout the benefits of mindfulness meditation for practicing lawyers, particularly in negotiation.[449] They contend that by adopting a non-judgmental perspective, mindfulness devotees "respond more appropriately to situations—and the thoughts, feelings, and bodily sensations that the situations elicit in us—rather than reacting in habitual ways."[450] Enthusiasts contend that Buddhist principles underlying mindfulness meditation and specific practice techniques can bring clarity of purpose and enhanced attention,[451] greater awareness and cognitive flexibility,[452] and the ability to make better choices.[453]

---

448. *Id.*

449. Leonard L. Riskin, *The Contemplative Lawyer: On the Potential Contributions of Mindfulness Meditation to Law Students, Lawyers, and their Clients*, 7 HARV. NEGOT. L. REV. 1 (2002). Riskin claims that mindfulness practice could make lawyers and law students "feel better and perform better at virtually any task" by reducing stress and improving the ability to concentrate. *Id.* at 46. He maintains that mindfulness meditation can help develop five emotional and social competencies of "emotional intelligence": self-awareness, self-regulation, motivation, empathy, and social skills. *Id.* at 47 (citing DANIEL GOLEMAN, EMOTIONAL INTELLIGENCE: WHY IT CAN MATTER MORE THAN IQ (1995)); *see also* Darshan Brach, *A Logic for the Magic of Mindful Negotiation*, 24 NEGOTIATION J. 25 (2008); Clark Freshman et al., *Adapting Meditation to Promote Negotiation Success: A Guide to Varieties and Scientific Support*, 7 HARV. NEGOT. L. REV. 67 (2002) [hereinafter Freshman et al., *Adapting Mediation*]; Clark Freshman, *After Basic Mindfulness Meditation: External Mindfulness, Emotional Truthfulness, and Lie Detection in Dispute Resolution*, 2006 J. DISP. RESOL. 511 [hereinafter Freshman, *After Basic Mindfulness Mediation*].

450. Riskin, *supra* note 449, at 29.

451. Brach, *supra* note 449, at 27–28.

452. Freshman et al., *Adapting Mediation*, *supra* note 449, at 74. Freshman et al. identify empirical support for professed benefits of mindfulness, citing psychological studies. *Id.* at 72–77. Research "neatly shows that both regular concentration and mindfulness meditation are associated with greater awareness." *Id.* at 74. Importantly, awareness is essential to changing behavior. *Id.* at 74 (citing John D. Teasdale et al., *Metacognitive Awareness and Prevention of Relapse in Depression: Empirical Evidence*, 70 J. CONSULTING & CLINICAL PSYCHOL. 275 (2002)). "Social science research also suggests another promising object for mindful negotiators involves emotions." *Id.* at 79. Freshman et al. observe that mindfulness of emotions can improve "mood awareness," allowing negotiators to understand what objects and thoughts induce positive mood and better negotiation results. *Id.* at 80.

453. Riskin, *supra* note 449, at 66.

Mindfulness meditation may help mediators attain greater bias reduction competency.[454] If the ability to listen is the mediator's stock in trade and mindfulness helps lawyers surmount barriers to careful listening, mindful mediators would be free from "distracting thoughts and emotions, 'personal agendas,' and bias and prejudice based on the speaker's appearance, ethnicity, gender, speech or manner."[455] Mediators trained in mindfulness would be more conscious of bias and stereotypes seeping into their thoughts and judgments. With that heightened awareness, they could call upon improved concentration to make better choices in the way they conduct their mediations.[456]

In summary, mediators have the ability to enhance internal neutrality by adopting explicit plans to reduce the application of stereotypes activated through encounters with parties and by replacing biased thoughts and reactions with non-prejudiced ones. Mediators must be aware of and acknowledge unconscious biases in order to garner the motivation to self-correct. A mediator's de-biasing action plan should include external and internal motivation to intervene with disputants in an egalitarian manner, attentiveness to prejudice-related discrepancies, and application of cognitive resources to reduce biased judgments and actions. By adopting individual "implementation intention" goals and strategies, mediators can attenuate bias. To encourage and facilitate these efforts, mediation programs should incorporate bias-reduction teaching techniques, make bias and prejudice reduction a robust part of the curriculum, and develop protocols that stress self-awareness, self-monitoring, and self-correction. Practices that sharpen a mediator's awareness, listening skills, and concentration (such as mindfulness meditation) may help mediators attain freedom from bias and prejudice.

---

454. *See* Rock, *supra* note 50.

455. Riskin, *supra* note 449, at 50.

456. *See* Brach, *supra* note 449, at 28 (arguing that mindfulness techniques may enhance "capacity to focus and sustain our attention consciously so that we can make the choices that serve our truest purposes").

CONCLUSION

Extensive research and analysis related to mediator behavior, the dynamics of the mediation process, and the science of implicit social cognition reveal a huge gap between the vision of mediator neutrality and the realities of biased mediator thoughts and actions. Well-meaning mediators who espouse egalitarian views need more than a "wish and a prayer" to actualize non-biased feelings, behaviors, and judgments. When confronted with scientific findings and empirical evidence, mediation professionals must concede that the requirements for eliminating racial, gender, and other types of bias in mediation have not been met.

I present the small claims mediation scenario as an example of a situation in which no one refers to race but "race [is] speaking *sotto voce*."[457] These types of cases are can be instructive because they "reveal how profoundly issues of difference have permeated the unconscious as well as the consciousness of people in our society."[458] Reflecting on such a case, Gadlin muses, "At times I feel so conscious of the way my response to peoples' stories and interventions in their conflicts is infiltrated by my own racial/ethnic/gender identity."[459] Mediation practice would be substantially improved if all mediators attained an equally critical self-consciousness.[460]

My goal in writing this Article is to challenge mediation teachers, trainers, and practitioners to admit to impartiality shortcomings and undertake concrete measures to alter the way we think and act. Defining what it means to be racially unbiased also presents difficulties. Many people think that being unbiased means they do not "see" race, gender, or ethnicity. People claim to be "color-blind," viewing this as the achievement of a non-prejudiced state of mind. "According to the most straightforward account, to be racially unbiased would require one to accord race no more significance than,

---

457. Gadlin, *supra* note 442, at 34.
458. *Id.*
459. *Id.*
460. *Id.*

say, eye or hair color, and to act as though one does not notice race."[461] But

> social practices and legal rules permit, indeed encourage, some species of race consciousness that virtually no one views as morally objectionable. Identifying racial bias, then, must entail deciding that some forms of race consciousness are more, or less, morally objectionable than others, a determination with respect to which reasonable minds may differ.[462]

Race has been such a pervasive and salient feature in our history and current society that everyone is subject to race consciousness.[463] The pervasiveness of implicit bias opens a new route to discussions of bias prevention and mitigation. The existence of unconscious bias does not necessarily mean that people with egalitarian beliefs are racists or liars.[464] Having discriminatory thoughts does not mean a low-prejudiced person endorses the belief; rather, it is an indication of the vigor of well-learned cultural stereotypes.[465]

Along the same lines, the operation of stereotypes need not be illustrative of a person's "moral failure." Gary Blasi points out that moralizing strategies are ineffective in combating stereotypes and prejudice.[466] He contends that "the science [of implicit social cognition] demonstrates in many ways that there is unlikely to be such a thing as a *non*racialized setting in the United States, if we include the various ways in which race operates indirectly."[467] He urges legal scholars and advocates to become knowledgeable about research on cognitive science and social psychology so they may overcome their own biases.[468] I call on mediators to do the same, lest

---

461. R. Richard Banks et al., *Discrimination and Implicit Bias in a Racially Unequal Society*, 94 CALIF. L. REV. 1169, 1171 (2006) (examining race consciousness in the criminal justice system).
462. *Id.*
463. *Id.* at 1184.
464. Armour, *supra* note 144, at 18–21.
465. *Id.* at 20.
466. Gary Blasi, *Advocacy Against the Stereotype: Lessons from Cognitive Social Psychology*, 49 UCLA L. REV. 1241, 1271 (2002).
467. *Id.* at 1273.
468. *Id.*

we be well-meaning but ineffective actors in the struggle to eliminate bias in mediation.

In a study of interracial tension, Patricia Devine and Kristin Vasquez considered the problem that the good intentions of low-prejudiced people

> are useful only if they are accurately interpreted by the target of those intentions. Intentions cannot be seen and must be inferred from behavior. This could be a problem if, for example, minority group members rely on the types of nonverbal behaviors that do not distinguish between anxiety and hostility.[469]

The authors note the potential for miscommunication that could escalate rather than alleviate tension.[470] They suggest that "the single most important problem facing us over time is that we are afraid to communicate."[471] They provoke with questions:

> What if we gave up the pretense that we 'should know what to do'? What if we admitted ignorance when it exists and confessed our desire to learn and understand? . . . But this approach may be a better starting point for alleviating tension than trying to fake it through the interaction and worrying the whole time about what we're doing wrong.[472]

The veneer of neutrality is stripped away by research findings that show convincingly that mediators fall far short of the ethical duty to treat parties impartially and without bias. Under current conditions, we are failing to meet our articulated goals and the expectations of the parties. Surely, it is naïve to think we can completely eliminate bias in mediation. It is equally certain that nondiscrimination in mediation is attainable only with more deliberate, informed, and self-conscious practices by mediators.

---

469. Patricia G. Devine & Kristin A. Vasquez, *The Rocky Road to Positive Intergroup Relations*, in CONFRONTING RACISM: THE PROBLEM AND THE RESPONSE, *supra* note 244, at 234, 261.
470. *Id.*
471. *Id.* at 262.
472. *Id.*

The Media, A Polarized America & ADR Tools to Enhance Understanding of Perspectives

Ginsey Varghese Kramarczyk*

I. INTRODUCTION

In today's political climate, the U.S. media reports to a disillusioned and polarized American audience and is driven largely by news outlets capitalizing on "conflict" to determine "newsworthiness."[1] With the additional preeminence of social media and algorithms promoting self-reinforcing media, each person's news appetite is limited by confirmation bias.[2]

The media's influence and ability to enhance the public's understanding of varying perspectives is important to promote civil discourse with "the other side."[3] The media has the unique "power to identify, name, and shape issues," and serves as the "gatekeeper" that decides what is "news" and *how* it will be spun.[4] The media mediates information to the public, and its role should thus be considered and guarded purposefully.[5]

"Media," as used in this paper, refers to "news media" or the "press" that delivers news to the public and practices journalism via various mediums—television, radio, magazines, and newspaper, both print and online.[6] More

---

* B.A., Oral Roberts University, 2010; M.S., Oklahoma State University, 2013; J.D. Candidate, Pepperdine University School of Law, 2019. The author reserves the right to depart from these views in the future.

[1] *Political Polarization in the American Public*, PEW RESEARCH CENTER (June 12, 2014), http://www.people-press.org/2014/06/12/political-polarization-in-the-american-public/ [hereinafter Pew, *Polarization in America*]; Carol Pauli, *News Media as Mediators*, 8 CARDOZO J. CONFLICT RESOL. 717, 717 (2007).

[2] *See* Maeve Duggan & Aaron Smith, *The Political Environment on Social Media*, PEW RESEARCH CENTER (Oct. 25, 2016), http://www.pewinternet.org/2016/10/25/the-political-environment-on-social-media/.

[3] *15.7 Media Influence on Laws and Government*, M LIBRARIES, http://open.lib.umn.edu/mediaandculture/chapter/15-7-media-influence-on-laws-and-government/ (last visited Sept. 29, 2018); *see also* DIANA C. MUTZ, HEARING THE OTHER SIDE: DELIBERATE VERSUS PARTICIPATORY DEMOCRACY 125 (2006) (discussing the consequences of interaction between people who have different political views).

[4] Lorie M. Graham, *A Right to Media?*, 41 COLUM. HUM. RTS. L. REV. 429, 429 (2010); *see also* Pamela J. Shoemaker, *News and Newsworthiness: A Commentary*, 31 EUR. J. COMM. RES. 105, 109 (2006).

[5] BILL KOVACH & TOM ROSENSTIEL, THE ELEMENTS OF JOURNALISM: WHAT NEWSPEOPLE SHOULD KNOW AND THE PUBLIC SHOULD EXPECT 166 (Three Rivers Press ed., 2d ed. 2007).

[6] "Media," "Journalist," and "Press" will be used interchangeably throughout the article. STEPHANIE CRAFT & CHARLES N. DAVIS, PRINCIPLES OF AMERICAN JOURNALISM: AN INTRODUCTION 11 (Erica Wetter ed., 2013).

specifically, "news media" refers to "organizations in the business of gathering and disseminating news, and to reporters, editors, and others who are professional news gatherers and disseminators, primarily at the national level differentiat[ing] the news function from the entertainment function."[7]

This article will survey: (1) the intended role of the media in a democracy; (2) the current polarized political climate in the United States; (3) the challenges facing the twenty-first century with the growth of technology, cable news, and online platforms; (4) the media's role in perpetuating conflict; and (5) propose that media professionals use Alternative Dispute Resolution (ADR) tools and processes to increase the public's understanding of differing perspectives in our conflict-laden political discourse.

A. *Potential Objections*

The First Amendment states, "Congress shall make no law . . . abridging the freedom of speech, or of the press."[8] Proposing a regulatory framework for the media or limiting any form of speech is not in the scope of this article. This article will survey the intended role and responsibilities of the media, analyze current contextual challenges, and provide ADR insights to empower news media professionals who are cognizant of their responsibility to facilitate a public forum to better serve the citizenry in today's polarized political climate.

Movements in the media toward using ADR techniques, such as "peace journalism" and "civic journalism," have faced resistance due to conflicts of interest for reporters.[9] Journalists are expected to maintain a sense of detachment and neutrality to ensure "factual reporting" and ADR "problem-solving" approaches do not pose the risks of agenda pushing.[10] As such, this article will not suggest that "problem-solving" techniques should take priority over "factual reporting."

Predominant today, divisive "reporting" is fueled by interests in higher ratings and financial pressures, likely incentivizing journalistic impropriety.[11] The press was created to serve as "responsible members of the community with a full stake in public life and thus 'concerned' with whether genuine citizen deliberation occurs when needed, and whether communities come to

---

[7] CRAFT & DAVIS, *supra* note 6, at 92.

[8] U.S. CONST. amend. I.

[9] Carol Pauli, *Transforming News: How Mediation Principles Can Depolarize Public Talk*, 15 PEPP. DISP. RESOL. L.J. 85, 100 (2015).

[10] *See generally* EUROPEAN CENTRE FOR CONFLICT PREVENTION, THE POWER OF THE MEDIA: A HANDBOOK FOR PEACEBUILDERS (2003); Annabel McGoldrick & Jake Lynch, *Peace Journalism: What Is It, How to Do It?*, REPORTING THE WORLD (2000), https://www.transcend.org/tri/downloads/McGoldrick_Lynch_Peace-Journalism.pdf.

[11] Jürgen Krönig, *A Crisis in the Fourth Estate*, THE GUARDIAN (Aug. 16, 2004), https://www.theguardian.com/media/2004/aug/16/mondaymediasection.politicsandthemedia (concluding that the media are driven by commercial and market pressures).

grips with their problems."[12] This article submits that journalists should have an inherent interest in sound reporting and the psychology of conflict and communications.   This entails understanding how "packaging" and "messaging" ultimately impact the audience,[13] but nevertheless the article does not propose regulating journalistic freedom in any manner.   ADR techniques would only serve to supplement newspersons' toolkits as they inform and educate the public on relevant issues, while providing for the differing perspectives of fellow citizens.[14]

II. MEDIA IN PUBLIC LIFE

A. *Media as the Fourth Branch of Government*

Many scholars regard the media as the "fourth branch" of government, fulfilling a necessary "check" function in a democracy over the standard three branches of government: legislative, executive, and judicial.[15] The First Amendment captures this fundamental ideal by providing for the "freedom of press" as distinct from the freedom of speech.[16] The founders explicitly created *a free press* to prevent tyrannical rule.[17] The press would serve as a protector against "abuse of official power" and "an educator of the people" by informing the people about government action and policies, as well as possible repercussions and options.[18] A free press, by design, would serve as a conduit between governmental institutions and the public: requiring accountability in government actions.[19]

NY Times Co. v. U.S. states this proposition clearly: "The Founding Fathers gave the free press the protection it must have to fulfill its essential

---

[12] Tanni Haas & Linda Steiner, *Public Journalism as a Journalism of Publics*, 2 SAGE JOURNALS 123, 132 (2001), http://journals.sagepub.com/doi/pdf/10.1177/146488490100200202.

[13] Peter du Toit, *Conflict Sensitive Reporting: A Toolbox For Journalists*, RHODES UNIVERSITY 43–44 (Mar. 2012), https://www.internews.org/sites/default/files/resources/ConflictSensitiveReporting-Peter_du_Toit_2012-03.pdf (providing guidance on how journalists can use the psychology of conflict and communication in reporting).

[14] *See* Rukhsana Aslam, *The Role of Media in Conflict: Integrating Peace Journalism in the Journalism Curriculum*, 8 (Nov. 4, 2014) (unpublished Ph.D. thesis, Auckland University of Technology) http://aut.researchgateway.ac.nz/bitstream/handle/10292/7908/AslamR.pdf?sequence=3 (arguing for the implementation of conflict analysis, resolution, and development in journalistic education).

[15] This article holds the premise that media is the unofficial fourth branch of government. Rachel Luberda, *The Fourth Branch of Government: Evaluating the Media's Role in Overseeing the Independent Judiciary*, 22 NOTRE DAME J.L. ETHICS & PUB. POL'Y 507, 508 (2008); William T. Coleman, Jr., *A Free Press: The Need to Ensure an Unfettered Check on Democratic Government Between Elections,* 59 TUL. REV. 243, 243–44 (1984).

[16] Coleman, *supra* note 15, at 243.

[17] *Id.* at 243–44.

[18] *Id.* at 247, 252.

[19] *Id.* at 244, 252.

role in our democracy. The press was to serve the governed, not the governors."[20] Justice Stewart highlights in a dissenting opinion that because "[e]nlightened choice by an informed citizenry is the basic ideal upon which an open society is premised . . . , the [independent] press . . . is a precondition of government."[21] In a democracy where government is to be by the consent of the governed, the media is *how* the public learns about the government.[22] Therefore, the media's role cannot be overlooked, downplayed, or misunderstood. The media controls the news it reports to the public and directly influences public opinion: holding a crucial, tremendous power in facilitating public life.[23]

B.   *Media's Responsibility to the Citizenry*

Media and the field of journalism have evolved over the years with the prevalence of Facebook, personal blogs, and Google, alongside major changes in culture, politics, and technology.[24] Part III of the article will discuss the challenges of twenty-first century technology and social media influences. Despite the recent evolution, the objectives of news media remain unaffected: "[T]o provide people with the information they need to be free and self-governing . . . amplifying the conversation of people themselves."[25] The media maintains a responsibility to provide independent, reliable, and comprehensive information for citizens to formulate opinions and to properly engage in the democratic process.

However, the difficulty is that journalism *needs* to make money in order to survive, and conflict sells.[26] Kovach, a prominent voice in journalism, outlines the ten principles of journalism for news people amidst the changing environment:

1) Journalism's first obligation is to truth. 2) Its first loyalty is to citizens. 3) Its essence is a discipline of verification. 4) Its practitioners must maintain

---

[20] *New York Times Co. v. United States*, 403 U.S. 713, 717 (1971).

[21] *See* Christa Corrine McLintock, *The Destruction of Media Diversity, or: How the FCC Learned to Stop Regulating and Love Corporate Dominated Media*, 22 J. MARSHALL J. COMPUTER & INFO. L. 569, 571 n.21 (2004) (quoting Branzburg v. Hayes, 408 U.S. 665, 727 (1972)) (Stewart, J. dissenting).

[22] RICHARD DAVIS, THE PRESS AND AMERICAN POLITICS (Beth Mejia ed., Prentice Hall 3rd ed. 2001); *see* KOVACH & ROSENSTIEL, *supra* note 5, at 11 (noting how journalism evolved to create democracy in countries like Poland and Czech Republic—"journalism was for citizenship. Journalism was for democracy.").

[23] KOVACH & ROSENSTIEL, *supra* note 5, at 11–12 ("The principles and purpose of journalism are defined by something more basic: the function news plays in the lives of people."); Diana C. Mutz & Paul S. Martin, *Facilitating Communication Across Lines of Political Difference: The Role of Mass Media*, 95 AM. POL. SCI. REV. 97, 109–10 (2001); *see* William L. Rivers, *The Media as Shadow Government*, in IMPACT OF MASS MEDIA: CURRENT ISSUES 279, 282 (1985) (emphasizing that the media sets the agenda for public discussion by controlling access to what the public reads).

[24] KOVACH & ROSENSTIEL, *supra* note 5, at ix.

[25] *Id.* at 3, 5–6, 11–12.

[26] Corporate, financial, and rating pressures will be discussed in Part IV.

independence from those they cover. 5) It must serve as an independent monitor of power. 6) It must provide a forum for public criticism and compromise. 7) It must strive to make the significant interesting and relevant. 8) It must keep the news comprehensive and in proportion. 9) Its practitioners have an obligation to exercise their personal conscience. 10) Citizens, too, have rights and responsibilities when it comes to the news.[27]

The importance of such tenants cannot be understated considering the unique challenges of the political climate today. Kovach warned that independent news could be overrun by rumors and self-interested commercialism posing as news.[28] He added, "[i]f that occurs, we lose the press as an independent institution, free to monitor the other powerful forces and institutions in society."[29] This is the situation in America today. The public is steadily losing trust in the press as independent and unbiased, with the prevalence of rhetoric about "fake news" and general cynicism dominating political discourse.[30]

III. POLARIZATION IN THE UNITED STATES

Polarization occurs when the judgment of individuals in a group regarding an issue becomes more extreme after discussing that issue with others in the group.[31] While there is markedly increased hostility in the U.S. political environment, there is disagreement amongst researchers as to whether there is a "polarization of ideologies" in the United States.[32] One group of researchers reason that Americans have become more extreme in their policy beliefs, leading to a strong "ideological polarization," while another group of researchers contend that most Americans' views have remained mostly stable over time.[33] Despite mixed findings regarding

---

[27] KOVACH & ROSENSTIEL, *supra* note 5, at 5–6.

[28] *Id.* at 6.

[29] *Id.* at 6.

[30] 57% of people in the US believe the media does a poor job of reporting on politics fairly. Amy Mitchell, Katie Simmons, Katherine Eva Matsa & Laura Silver, Pew Research Center, *Publics Globally Wanted Unbiased News Coverage But Are Divided Whether Their News Media Delivers* (Jan. 2018); Michael Barthel & Amy Mitchell, Pew Research Center, *Americans' Attitudes About the News Media Deeply Divided Along Partisan Lines* (May 2017); Art Swift, *Americans' Trust in Mass Media Sinks to New Low*, GALLUP NEWS (Sept. 14, 2016), http://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx; Pauli, *News Media As Mediators, supra* note 1, at 717.

[31] DOUGLAS KENRICK, STEVEN L. NEUBERG & ROBERT B. CIALDINI, SOCIAL PSYCHOLOGY: GOALS IN INTERACTIONS 392 (6th ed. 2015).

[32] Matthew Gentzkow, *Polarization in 2016*, STAN. U. 1, 2 (2016), https://web.stanford.edu/~gentzkow/research/PolarizationIn2016.pdf.

[33] *See generally* M. P. Fiorina & S. A. Abrams, *Polarization in the American Public*, 11 ANN. REV. OF POL. SCI, 563, 563–88 (2008); Alan I. Abramowitz & Kyle L. Saunders, *Is Polarization a Myth?*, 70(2) J. OF POL, 542, 542–55 (2008); Edward L. Glaeser, & Bryce A. Ward, *Myths and Realities of American Political Geography*, 20 J. OF ECON. PERSP., 119, 119–44 (2006).

ideological polarization, researchers agree that Americans are experiencing "affective polarization."[34]

"Affective polarization" can be defined as the tendency of people to view "in-group" members positively while viewing "out-group" members negatively.[35] It is the inclination of people to assess the opposing partisans negatively and co-partisans positively.[36] As a result, Americans feel distant from members of the opposite party and even dislike those on the other side of the political spectrum.[37] In the current political climate, many Americans who identify with a particular party hold extreme biases against the opposing party members for simply identifying as a "Democrat" or "Republican," even if associated policies or ideologies are not, in reality, extreme.

This type of behavior, while not suggestive of ideological polarization, highlights that many voters have become polarized by "the mere act of identifying with a political party . . . [triggering] negative evaluations of the opposition, [with] exposure to prolonged media-based campaigns only reinforc[ing] these predispositions."[38] Seyle and Newman, both prominent social psychologists, speculate that the majority of America remains ideologically moderate, but the "red" versus "blue" emphasis drowns moderate voices and divides voters into two opposing camps represented by the loudest extremists on both sides.[39] These distortions, according to experts in social psychology, "increase group conflict, decrease intergroup contact, and contribute to many problems in social interaction across groups."[40]

The public sphere is becoming increasingly polarized. A recent 2017 Pew survey revealed that on the eve of Trump's inauguration, 86% of Americans believed that the United States was more politically divided than in the past.[41] Only 46% of Americans held such conviction when Barack Obama became President of the United States in 2009.[42] This statistic reflects that 40% of the American public has recognized this shift within the past eight years.[43]

A 2014 Pew study on polarization in America revealed that politically active Americans, 27% of Democrats and 36% of Republicans, increasingly saw their opponents as "so misguided that they threaten the nation's

---

[34] Shanto Iyengar & Sean J. Westwood, *Fear and Loathing Across Party Lines: New Evidence on Group Polarization*, 59(3) AM. J. OF POL. SCI., 690, 690–707 (2015).

[35] *Id.* at 691.

[36] *Id.* at 691.

[37] *Id.* at 691.

[38] Shanto Iyengar, Gaurav Sood & Yphtach Lelkes, *Affect, Not Ideology: A Social Identity Perspective on Polarization*, 76(3) PUB. OPINION Q. 405, 407 (2012).

[39] D.C. Seyle & M.L. Newman, *A House Divided? The Psychology of Red and Blue America*, 61(6) AM. PSYCHOLOGIST 571–80 (2006).

[40] *Id.* at 574.

[41] Pew Research Center, *On Eve of Inauguration, Americans Expect Nation's Deep Political Divisions to Persist* (Jan. 2017), http://www.people-press.org/2017/01/19/on-eve-of-inauguration-americans-expect-nations-deep-political-divisions-to-persist/ [hereinafter Pew, *On the Eve of Inauguration*].

[42] *Id.*

[43] *Id.*

wellbeing."[44] Nearly half of politically active Americans, deemed as "active" because of voting patterns and attention to political news, regarded the other side as a danger to the nation.[45] It flows naturally that more than one-fifth of the respondents in the same survey indicated they would be displeased if one of their children married someone of the opposite party.[46] Compared to 1960, where very few people (5%) indicated this sentiment, the affective polarization, or rather "personalization," of politics is apparent.[47] Polarization appears to be getting worse under Trump's presidency.

Two key shifts have occurred within each party: (1) people's connection with their party is deeper and (2) consistency of people's views has increased.[48] The first shift is apparent in the numbers, but the shocking change is that politics have become incredibly personal.[49] Party affiliation has evolved to be a part of our self-concept in profound ways.[50] This connection of party and "self" extends to how individuals judge politics and new information.[51] People do not view the other party as made up of competent individuals who espouse a different viewpoint or hold a different value system.[52] Instead, the other side is perceived as selfish and stupid, with intolerable views or, at the very least, precarious motives.[53] The other side is seen as a threat to the nation.[54]

The negative sentiment is captured perfectly in party favorability ratings presented in an October 2017 Pew report.[55] It reflects that "about eight-in-ten Democrats and Democratic-leaning independents (81%) have an unfavorable opinion of the Republican party."[56] This includes 44% holding a "very unfavorable view" of the Republicans.[57] The data is the same amongst Republicans: 81% view Democrats unfavorably, with 45% holding a "very unfavorable view."[58] Twenty years ago, a little over half of Democrats (57%) viewed the Republican party unfavorably and only 16% very unfavorably,

---

[44] Pew, *Polarization in America, supra* note 1, at 8, 11; Pauli, *Transforming News: How Mediation Principles Can Depolarize Public Talk, supra* note 9; Gentzkow, *supra* note 29, at 15.

[45] Pew, *Polarization in America, supra* note 1, at 8, 11; Gentzkow, *supra* note 29, at 15.

[46] Pew, *Polarization in America, supra* note 1, at 12.

[47] *Id.* at 12.

[48] *Id.* at 11; Gentzkow, *supra* note 32, at 17.

[49] *See* Gentzkow, *supra* note 32, at 1.

[50] Alexander George Theodoridis, *The Hyper-Polarization of America*, Sci. Am. (Nov. 7, 2016), https://blogs.scientificamerican.com/guest-blog/the-hyper-polarization-of-america.

[51] Theodoridis, *supra* note 50.

[52] Gentzkow, *supra* note 32, at 17.

[53] Gentzkow, *supra* note 32, at 17; *see* Pew Research Center, *The Partisan Divide on Political Values Grows Even Wider* (Oct. 2017) [hereinafter Pew, *Divide on Political Values Grows Wider*].

[54] Pew, *Polarization in America, supra* note 1, at 8, 11; Theodoridis, *supra* note 50.

[55] *See* Pew, *Divide on Political Values Grows Wider, supra* note 53, at 65–66.

[56] *Id.* at 66.

[57] *Id.*

[58] *Id.*

134   *Pepperdine Dispute Resolution Law Journal*   [Vol. 19:1

while 68% of Republicans viewed the Democratic Party negatively and only 17% very unfavorably.[59] Today, negative opinions are commonplace and more strongly felt than in the past.

On the second point, Americans divide more easily into two separate camps of Republican and Democratic views.[60] This considers independents, who actually outnumber both Republicans and Democrats today, because independents ideologically lean to one party or another.[61] Americans might not be extreme in their overall ideology, but they hold extreme positions on specific issues.[62] In the past, it was more common to find a conservative democratic or liberal Republican.[63] Today, there are fewer people who hold such hybrid views—liberal on some issues and conservative in other areas.[64] Apparently, an individual's choice of presidential candidate increasingly represents how the person regards a spectrum of issues: gun control, abortion, welfare, taxes, racial issues, Islam, the environment, and so on.[65]

According to the ten indices that Pew Research Center have measured via the same questions since 1994, the average partisan gap has increased from fifteen percentage points to thirty-six percentage points on each issue.[66] Never before has the disparity been this large. For example, in 2011, "twice as many Democrats as Republicans said the government should do more for the needy (54% vs. 25%). Today, nearly three times as many Democrats as Republicans" hold this view (71% vs. 24%).[67] The widening partisan gap exists on other issues like gun control and racial discrimination as well.[68] It should be noted, however, as a result of greater acceptance of homosexuality and more favorable views on immigrants (as less of a burden on the country at least), the center has shifted in a liberal direction overall.[69]

Evidence indicates that polarization is a growing concern in America today.[70] There is disagreement about how far apart Americans are on the issues, but the divide is undeniably more personal.[71] People no longer politely disagree because they believe the other side must be stopped.[72] The news

---

[59] *Id.*

[60] Gentzkow, *supra* note 32, at 17; Pew, *Divide on Political Values Grows Wider, supra* note 53, at 11–12.

[61] Pew Research Center, *Partisanship and Political Animosity in 2016,* 6 (June 2016) [hereinafter, Pew, *Political Animosity*].

[62] Gentzkow, *supra* note 32, at 17.

[63] *Id.* at 17.

[64] *Id.*

[65] Pew, *Divide on Political Values Grows Wider, supra* note 53, at 3, 7–12.

[66] *Id.* at 3, 12.

[67] *Id.* at 2.

[68] *Id.* at 2.

[69] *Id.* at 11.

[70] Gentzkow, *supra* note 32, at 20.

[71] *Id.*

[72] *Id.*; Pew, *Political Animosity, supra* note 61, at 6.

media, especially the growth of partisan cable news and confirmation bias perpetuated by technology and online platforms, likely played a role in this outcome, as discussed in the next section. [73]

IV. CHALLENGES OF THE TWENTY-FIRST CENTURY

A. *Rise of Technology & Social Media*

The way Americans consume the news has changed with the rise of technology and social media platforms. Nearly all adults in America (99%) own at least one electronic device (including a television).[74] According to a Pew study in 2015, 68% of U.S. adults own a smartphone, 73% own laptops or desktop computers, and 45% own tablets (up from 4% in 2010).[75] The American Psychological Association estimated even higher numbers in 2016, with 86% computer ownership, 74% smartphone ownership, and 55% tablet ownership.[76] Among young adults between the ages of eighteen and twenty-nine in both surveys, ownership of smartphones with internet access reached almost 90% of the population.[77]

---

[73] Gentzkow, *supra* note 32, at 20.

[74] Am. Psychological Ass'n, *Stress in America: Coping with Change, Stress in American Survey,* 1 (2017), https://www.apa.org/news/press/releases/stress/2017/technology-social-media.pdf.

[75] Pew Research Center, *Technology Device Ownership* 3 (Oct. 2015).

[76] Am. Psychological Ass'n, *supra* note 74, at 1.

[77] Pew, *Technology Device Ownership,* supra note 75, at 3; Am. Psychological Ass'n, *supra* note 74, at 1.

The number of American adults on social media increased to 65% in 2015, compared to 7% in 2005.[78] Among young adults, social media is approaching market saturation with over 90% young adult social media users reported in 2015.[79] Facebook and Instagram today celebrate over 2 billion combined monthly users.[80] Compared to figures in the last decade, these are substantial increases in electronic device ownership and social media adoption, both of which lead to an increase in news media exposure.[81]

B. *Trends in Technology Usage*

On a typical day, 44% of Americans report constantly checking their social media.[82] An American adult generally looks at their phone once every twelve minutes, or eighty times a day,[83] and, collectively, Americans check their phones over eighty billion times a day.[84] A mobile consumer survey conducted by Deloitte found that "reading the news" was ranked as the "most regularly used daily activity" in smartphone additional uses.[85]

The modern audience consumes news on a screen. Television remains dominant with 57% of Americans relying on television-based news and 38% relying on online news via websites, apps, social media or all three.[86] On the other hand, the portion of American adults who have *ever* accessed news on a mobile device increased from 54% in 2013, to 72% in 2016, to 85% in 2017.[87] While a majority of Americans, especially the older demographics of ages fifty to sixty four (72%) and over the age of sixty-five (85%), chiefly depend on and prefer television, the impact of online platforms cannot be discounted because adults under age forty-nine (94%) increasingly rely on news via their mobile devices and online platforms.[88] Older adults are transitioning to smartphones as well, with about two-thirds of older Americans (67%), ages

---

[78] Andrew Perrin, *Social Media Usage: 2005–2015*, PEW RESEARCH CENTER, (Oct. 8, 2015), http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/.

[79] *Id.* at 4–5.

[80] Am. Psychological Ass'n, *supra* note 74, at 1.

[81] Perrin, *supra* note 78; Pew, *Technology Device Ownership*, *supra* note 75, at 3; *see* Lisa Eadicicco, *Americans Check Their Phones 8 Billion Times a Day*, TIME (Dec. 15, 2015) http://time.com/4147614/smartphone-usage-us-2015.

[82] Am. Psychological Ass'n, *supra* note 74, at 4.

[83] SWNS, *Americans Check Their Phones 80 Times a Day: Study*, NY POST (Nov. 8, 2017, 4:08 PM), https://nypost.com/2017/11/08/americans-check-their-phones-80-times-a-day-study.

[84] Eadicicco, *supra* note 81.

[85] *Global Mobile Consumer Survey: US Edition*, DELOITTE, https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-survey-us-edition.html; *See* Eadicicco, *supra* note 81.

[86] Amy Mitchell, Jeffrey Gottfried, Michael Barthel & Elisa Shearer, *Modern News Consumer*, PEW RESEARCH CENTER (July 7, 2016), http://www.journalism.org/2016/07/07/pathways-to-news/.

[87] *Id.* at 5; Kristine Lu, *Growth in mobile news use driven by older adults*, PEW RESEARCH CENTER (Jun. 12, 2017), http://www.pewresearch.org/fact-tank/2017/06/12/growth-in-mobile-news-use-driven-by-older-adults.

[88] Mitchell et al., *supra* note 86, at 5 (finding that 50% of 18–29 year olds and 49% of 30–49 year olds turn to online platforms for news).

sixty-five and over using a mobile device for news – this represents an increase of 24% since 2016 alone.[89] The growth in the usage of devices— TV, computer, mobile phones, and tablets—to consume the news, raises concerns about the impact on the American public. Does the news media play a role in the polarization and personalization of politics in America today? If so, what can be done?

C. *Fragmented Media & Echo Chambers*

Four decades ago, most Americans watched daily news from one of three newscasts in a standard and homogeneous "point-counterpoint" style. [90] However, today's news media is highly fragmented with 24-hour cable news, newspapers, bloggers, online platforms, talk-radio, and other outlets simultaneously competing for the public's attention.[91] With a plethora of news outlets and platforms at the public's fingertips via their devices, the news media must compete to stay relevant.

People generally feel strongly about their positions or policy preferences and seek out information consistent with their paradigms.[92] Kovach warned that "the ability of almost anybody to produce and disseminate text, video, and audio . . . [increases] user demand to personalize the content they consume from other sources."[93] This type of selective exposure produces the "echo chamber" effect where news reinforces one's beliefs and attitudes; this creates what psychologists call "confirmation bias" — the tendency of people to embrace information that supports one's beliefs and attitudes and rejects contradictory messages.[94]

To demonstrate, if a person believes people who live on the coasts are liberal, confirmation bias means that a person gravitates towards evidence that the belief held is true. Therefore, when that person encounters one individual who leans liberal who lives on the coast, the person is likely to believe the data is correct because it reinforces an existing belief. However, if the person encounters a conservative who lives on the coast, the individual easily dismisses the information as a fluke. Confirmation bias occurs when a person encounters information that supports an existing belief, and that person subconsciously place greater importance on that "evidence."[95] The bias

---

[89] Lu, *supra* note 87 (noting that those between the ages of 50–64, 79% now use mobile for news).

[90] Shanto Iyengar & Kyu S. Hahn, *Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use*, 50 J. COMM. 19, 20 (2009).

[91] *Id.* at 20.

[92] *Id.* at 20.

[93] KOVACH & ROSENSTIEL, *supra* note 5, at xi.

[94] Iyengar & Hahn, *supra* note 91, at 34; Elizabeth Kolbert, *Why Facts Don't Change our Minds*, NEW YORKER (Feb. 27, 2017), https://www.newyorker.com/magazine/2017/02/27/why-facts-dont-change-our-minds.

[95] *See* Kolbert, *supra* note 94.

explains why anti-vaccine narratives persist across the internet even though the link to autism has been repeatedly debunked.[96]

According to researchers, "[o]nce formed, . . . impressions are remarkably perseverant."[97] In a famous Stanford experiment, students were given two studies about capital punishment — one supporting its ability to deter crime while the other claimed it had no effect on crime.[98] Half of the selected participants supported capital punishment while the other half opposed it.[99] Despite the reality that both studies were fictitious, at the end of the experiment, each set of students found the data supporting their viewpoint to be compelling and deemed the opposing study unconvincing.[100] The experiment demonstrates the staying power of a belief that an individual holds. In a separate experiment, it was *revealed* to participants that information originally given was fabricated, but individuals failed to make appropriate revisions in their beliefs.[101] Researchers are fascinated by the confounding nature of confirmation bias in repeated experiments and how even a proven false impression can persist.[102]

In another experiment, people were asked to rate their understanding of how everyday items like toilets and zippers function.[103] Often, their rating of their true understanding reduced drastically after they were asked to write step-by-step explanations.[104] This phenomenon is labeled the "illusion of explanatory depth," where people believe that they know more than they do *and* if other individuals agree, their belief grows.[105]

How toilets flush is trivial compared to political questions that impact a community; yet, confirmation bias and the illusion of explanatory depth are at play. To illustrate, in a survey considering U.S. military intervention in a particular foreign country, the respondents who were least likely able to identify the country in question were most likely to favor military intervention.[106] This is particularly alarming.

To understand the "political brain," a neuroimaging study was conducted of a group of strong Republicans and strong Democrats.[107] In the study,

---

[96] David Grimes, *Echo Chambers Are Dangerous,* GUARDIAN (Dec. 4, 2017), https://www.theguardian.com/science/blog/2017/dec/04/echo-chambers-are-dangerous-we-must-try-to-break-free-of-our-online-bubbles; Luke E. Taylor, Amy L. Swerdfeger & Guy D. Eslick, *Vaccines are not Associated with Autism,* VACCINE 3623 (2014).

[97] Kolbert, *supra* note 94.

[98] *Id.*

[99] *Id.*

[100] *Id.*

[101] *Id.*

[102] *See id.*

[103] *Id.*

[104] *Id.*

[105] *Id.*

[106] *Id.*

[107] Michael Shermer, *The Political Brain,* SCI. AM. (July 1, 2006), https://www.scientificamerican.com/article/the-political-brain/?print=true.

subjects were critical of the opposing presidential candidate while repeatedly letting their own candidate off the hook.[108] This is how confirmation bias works. Most disturbing was that the MRI revealed that the dorsolateral prefrontal cortex, the part of the brain associated with *reasoning*, was inactive.[109] But the orbital frontal cortex, the part of the brain associated with processing emotions, and other parts of the brain associated with conflict resolution and moral accountability, the anterior cingulate and posterior cingulate respectively, were stimulated.[110] Once participants arrived at a conclusion with which they were emotionally comfortable, the ventral striatum, the reward and pleasure center, was found to be active.[111] Basically, people experience a rush of dopamine when processing information that supports their beliefs.[112]

The study emphasized two learnings: (1) partisans "twirl the cognitive kaleidoscope" to obtain the conclusions *they want*, and (2) brain circuits trigger rewards for these selective behaviors.[113] Many individuals' opinions frequently are baseless, and generally, "strong feelings about issues do not emerge from deep understanding" as these studies have shown.[114] Moreover, when people with baseless opinions agree, they reinforce one another while readily dismissing any contradictory evidence.[115] Subsequently, these groups feel empowered in their resistance or for "stick[ing] to [their] guns."[116] Ultimately, groups become oblivious to their tribalism, creating an "echo chamber" of confirmation bias. The key concern is whether the rise of partisan news and online algorithms perpetuate this "echo chamber" effect, and thereby bear responsibility for the growing polarization.

D. *Partisan News & Online Algorithms*

In our modern world, cable news is more partisan, online algorithms personalize information to cater to individual tastes, and people aggregate into communities of interest both socially and geographically.[117] People choose

---

[108] *Id.*

[109] *Id.*

[110] *Id.*

[111] *Id.*

[112] Kolbert, *supra* note 94.

[113] Shermer, *supra* note 107.

[114] Kolbert, *supra* note 94.

[115] *See id.*

[116] *See id.*; Shermer, *supra* note 107 (noting that the MRI study hypothesized that skepticism may be the antidote for the confirmation bias).

[117] *See* Kolbert, *supra* note 94; Levi Boxell, Matthew Gentzkow & Jesse Shapiro, *Is Media Driving Americans Apart?*, N.Y. TIMES (Dec. 6, 2017), https://www.nytimes.com/2017/12/06/opinion/is-media-driving-americans-apart.html; Am. Press. Inst., *Partisanship and News Behavior* (July 13, 2017), https://www.americanpressinstitute.org/publications/reports/survey-research/partisanship-news-behavior.

what to read or watch based on their own brand of politics, and as a result, the "media becomes background noise . . . [and] the journalism of affirmation [becomes] more appealing."[118]

As mentioned earlier, television is the medium by which a majority of Americans (57%) consume the news.[119] The American Press Institute (API) reported that cable news is the most popular news source among Americans, while broadcast news is a close second. [120] API's study found that Republicans are more likely to tune into Fox News and Democrats are more likely to rely on CNN and MSNBC. [121] A study of "large-scale data [pinpoints] cable television news as a major contributor to polarization."[122] This analysis is consistent with the growth in interparty hostility and the rise in polarization in groups like the elderly, who have "limited internet use but high rates of television viewing."[123]

On average, 57% of people in the U.S. believe the media does a poor job of reporting on politics fairly.[124] The debate between left and right-wing media bias spans several decades.[125] However, there is no agreement on what qualifies as media bias.[126] In fact, when two people with different viewpoints review identical content, even their perceptions of the "bias" differ.[127] On rudimentary review, the parties speak different languages, as seen in televised news rhetoric —Democrats' lingo routinely incorporates "undocumented workers" and the "estate tax" while Republicans use "illegal aliens" and "the death tax."[128]

Some researchers contend most news media outlets are centrist, and perhaps only slightly slanted, but provide a balanced mix of views—although some talk show radios and websites offer more extreme news and opinion.[129] Markus Prior, from Princeton University, argues that the culprit of increased partisan voting is not Fox News, but rather ESPN, HBO and other cable channels that lure the moderate voter away from the news. [130] Matthew Gentzkow, an economist and professor at Stanford University, agrees stating that the rise of television provided a medium to share political information, but it also offered Americans new ways to distract themselves in their free

---

[118] KOVACH & ROSENSTIEL, *supra* note 5, at 48–49.

[119] Mitchell, *supra* note 86, at 5; Boxell, *supra* note 117.

[120] Am. Press. Inst., *supra* note 117.

[121] *Id.*; Bharat N. Anand, *The U.S. Media's Problems Are Much Bigger than Fake News and Filter Bubbles*, HARV. BUS. REV. (Jan. 5, 2017).

[122] Boxell, *supra* note 117.

[123] *Id.*

[124] Mitchell, *supra* note 30, at 5; *see* Markus Prior, *Media and Political Polarization*, 16 ANN. REV. POL. SCI. 100, 101–27 (2013).

[125] Anand, *supra* note 121; *see* Am. Press. Inst., *supra* note 117.

[126] Anand, *supra* note 121.

[127] *Id.*

[128] *See* Boxell, *supra* note 117.

[129] Prior, *supra* note 124, at 103–04.

[130] *Id.* at 107.

time, crowding out the median's political engagement.[131] This suggests that the choice between "partisan or centrist" news may not be at issue if moderates avoid the news all together.[132] More study is needed on the moderate voter. For other individuals, selective exposure to news is simply a cognitive response to resist information that is inconsistent with their own views.[133]

Politically active voters, in particular, are less likely to be moved by counter-messaging because they collect a litany of information to contradict and neutralize disagreeable messaging.[134] Markus Prior agrees there is evidence that some individuals will gravitate toward like-minded news when given a choice between news reports, but other criteria will often supersede such partisan uniformity.[135] Researchers like Gerber, Karlan, Bergan, and others nevertheless maintain that media sources influence the public by the slant of a report, and more so, by the *choice* of news stories covered in a broadcast.[136]

Online platforms and social media also create division by their personalization of newsfeeds and search results. Among millennials, 61% claimed Facebook is a common source for political news.[137] Facebook, with almost half of U.S. adults getting their news from individual newsfeeds, is complicit in creating "filter bubbles" that personalize content using algorithms to accommodate individual subscribers' preferences.[138] Facebook's algorithm calculates one's preferences by analyzing a person's liked videos, recent conversations, frequent contacts, and content reviewed in order to provide more of the same content.[139]

This became apparent after the 2016 presidential election, when individual newsfeeds highlighted these personalized filters.[140] Identified Democrats *only* saw mourning of the presidential race while identified Republicans saw *only* celebration of the Trump victory despite having varied

---

[131] Matthew Gentzkow, *Television and Voter Turnout*, Q. J. ECON. 931, 970 (2006).

[132] Prior, *supra* note 124, at 120.

[133] *Id.*

[134] *Id.* at 108.

[135] *Id.* at 110.

[136] Alan S. Gerber, Dean Karlan & Daniel Bergan, *Does Media Matter?*, 1 AM. ECON. J. 35 (2009).

[137] Amy Mitchell, Jeffrey Gottfired & Katerina Eva Matsa, Pew Research Center, *Facebook Top Source for Political News Among Millenials*, PEW RESEARCH CENTER (June 1, 2015), www.journalism.org/2015/06/01/facebook-top-source-for-political-news-among-millennials/.

[138] Duggan & Smith, *supra* note 2; KOVACH & ROSENSTIEL, *supra* note 5, at 160; Sally Adee, *How Can Facebook and its Users Burst the 'Filter Bubble'?*, NEW SCIENTIST (Nov. 18, 2016), https://www.newscientist.com/article/2113246-how-can-facebook-and-its-users-burst-the-filter-bubble/.

[139] Selena Larson, *Facebook Shows You What You Want to See Post-Election*, CNNTECH (Nov. 9, 2016), http://money.cnn.com/2016/11/09/technology/filter-bubbles-facebook-election/index.html.

[140] *Id.*

friend groups.[141] While this concept and its impact raises questions on Facebook's civic responsibility, it is not the subject of this article.[142] This overt steering prevents individuals from evaluating various news sources and differing opinions objectively.[143] Personalized filters can also induce poor decisions based on faulty understanding.[144]

Moreover, aggregators of news clips and interviews from other sites have arisen as major "news source" players, including U.S. Uncut, Occupy Democrats, Addicting Info, Make America Great, and The Other 98%.[145] "Occupy Democrats, a far-left page popular with supporters of [] Democratic presidential candidate Bernie Sanders, has 3.8 million likes on its Facebook page. MSNBC . . . a mere 1.6 million."[146] This Facebook phenomenon underlines the missing dialogue between the opposing parties because of "echo chambers" and filter bubbles.[147]

The same data collection and configuration occurs with Netflix, Pandora, Google, and other social media platforms.[148] A thumbs-up on a site indicates one's preference and that data is captured by the algorithm.[149] If a person searches for almost anything on Google, the data is fed into an algorithm and begins to predict what an individual would like or would not like online.[150] This data collection directly impacts the news a person consumes because algorithms begin to predict what a person prefers, making conflicting and disagreeable voices disappear.[151]

Eli Pariser, the author of *The Filter Bubble: What the Internet is Hiding from You*, contends that many people are unaware of the voices that are missing.[152] Tailored recommendations categorically divide people into groups and limit their options and exposure.[153] Personalization of online news

---

[141] *Id.*

[142] Facebook is treated as a platform where news media is distributed by individuals or organizations. Michael Barthel, Elisa Shearer, Jeffery Gottfried & Amy Mitchell, *The Evolving Role of News on Twitter and Facebook*, PEW RESEARCH CENTER (July 14, 2015), http://www.journalism.org/2015/07/14/the-evolving-role-of-news-on-twitter-and-facebook/. This article does not treat Facebook as news media, and discussion on the issue is outside of the article's scope.

[143] Shermer, *supra* note 107, at 3.

[144] *Id.* at 3.

[145] Scott Bixby, *'The End of Trump': How Facebook Deepens Millennials' Confirmation Bias*, THE GUARDIAN (Oct. 1, 2016), https://www.theguardian.com/us-news/2016/oct/01/millennials-facebook-politics-bias-social-media.

[146] *Id.*

[147] *See* Natasha Singer, *The Trouble with the Echo Chamber Online*, N.Y. TIMES (May 28, 2011), http://www.nytimes.com/2011/05/29/technology/29stream.html.

[148] Twitter is different because posts are unfiltered, but individuals must sign up for a range of feeds to expand their exposure. Singer, *supra* note 147.

[149] Singer, *supra* note 147.

[150] *Id.*

[151] *See id.*

[152] *Id.*

[153] *Id.*

steers people to affirming content because the algorithms screen out news reports that individuals are likely to disagree with and creates a comfortable bubble of like-minded information.[154] The impact could be detrimental for a democracy because if people only receive affirming narratives, there is a lack of deliberation and understanding of varied interests in the public forum.[155]

On the other hand, some researchers argue that online filter bubbles are a myth because people do not receive their news from only one source.[156] In fact, individuals with particularly extreme views are more likely to consume a variety of news sources, including sites with conflicting ideology.[157] Pew Research reports that almost half of those who learned about the 2016 presidential election used five or more sources to gather their information.[158]

Like most issues, the polarization of America cannot be linked to any one change, but the convergence of many factors simultaneously.[159] This section discussed significant changes in the last decade, including the growth of devices, digital media, twenty-four-hour cable news, social media, online filter bubbles, and party tribalism. While not in the scope of this article, the lived experiences of Americans should not be disregarded—people are impacted by their self-sorting communities and socio-economic realities in a recovering economy. However, arguably, the most severe result of the many changes is that the digital world has undermined the business model that supports quality journalism.[160] The news media is struggling to stay relevant, which raises questions about the never-ending conflict narrative and the commercialization of the news as will be discussed in the next section.

V. MEDIA'S ROLE IN PERPETUATING CONFLICT

A. *Sensationalism & Commercialization of News*

When the public is more polarized, "news with an edge" garners market success.[161] At a White House press briefing about a successful economic plan with Ukraine during the Clinton presidency, one reporter expressed the ideology of mainstream journalism plainly: "Look, we have a rule here. 'No

---

[154] *Id.*; Gentzkow, *supra* note 32, at 18; Boxell, *supra* note 117.

[155] Singer, *supra* note 147.

[156] Matthew Gentzkow & Jesse Shapiro, *Ideological Segregation Online and Offline*, 126 Q. J. ECON. 1799 (2011).

[157] *Id.* at 1802.

[158] Jeffrey Gottfried, Michael Barthel, Elisa Shearer & Amy Mitchell, *The 2016 Presidential Campaign–A News Event That's Hard to Miss*, PEW RESEARCH CENTER (Feb. 4, 2016).

[159] Anand, *supra* note 121.

[160] Gentzkow, *supra* note 32, at 18.

[161] Iyengar & Hahn, *supra* note 94, at 19.

144   *Pepperdine Dispute Resolution Law Journal*   [Vol. 19:1

conflict, no story.'"[162] The media does not set out to be sensationalist, but its business model over the last few decades has anchored itself there.[163]

Since the 1990s, about six to eight media giants have controlled 90% of what we read, watch, or listen to.[164] This handful of corporations include Vivendi/Universal, AOL/Time Warner (CNN), The Walt Disney Co. (ABC), News Corporation (FOX), Viacom (CBS), General Electric (NBC), and Bertelsmann.[165] Traditional news outlets honored their civic responsibility to report on facts and provided important analysis by spending the money on field journalists to provide direct coverage on the ground.[166] Today, news channels bring in local experts and pundits to yell at each other on live broadcast because it attracts more viewers and advertisers, promoting sensational news over substantial news. [167] To the disappointment of committed journalists, sensationalist reporting and "entertainment" talk-shows or opinion news have proven to be highly popular according to the ratings.[168]

Big media conglomerates view large audiences as commodities for sale to advertisers.[169] The traditional model of news has been overrun by ad-revenue-driven news media and most Americans are prey to addicting sensationalist news or political disengagement.[170] Media professionals argue that *conflict captivates* the public,[171] just like when drivers slow down to observe a traffic accident even though it causes more unwanted traffic.[172] This style of conflict reporting indicates that people's news knowledge consists of episodic and fragmented accounts of dramatic moments,[173] which produces the "illusion of explanatory depth" discussed earlier in the article.[174] Instead of unwanted traffic, people suffer from a faulty understanding of surrounding causes and consequences on complex policy issues.[175]

---

[162] J. FALLOWS, BREAKING THE NEWS: HOW THE MEDIA UNDERMINE AMERICAN DEMOCRACY 164 (Pantheon 1996).

[163] Anand, *supra* note 121.

[164] Ashley Lutz, *These 6 Corporations Control 90% of the Media in America*, BUS. INSIDER (Jan. 14, 2012), http://www.businessinsider.com/these-6-corporations-control-90-of-the-media-in-america-2012-6.

[165] Alan B. Albarran & Terry Moellinger, *The Top Six Communication Industry Firms: Structure, Performance, and Strategy*, in MEDIA FIRMS: STRUCTURES, OPERATIONS, & PERFORMANCE, 102, 102–03 (2002).

[166] *See* Jacob W. Roberts, *The Tragedy of Media Sensationalism in America*, SOUTHERN CAL. INT'L REV. (May 20, 2014), http://scir.org/2014/05/the-tragedy-of-media-sensationalism-in-america.

[167] *Id.*

[168] *See id.*

[169] Phil Barker, *Large-Scale Communication*, BEYOND INTRACTABILITY (Mar. 2005), https://www.beyond intractability.org/essay/large-scale-communication.

[170] *See id.*; Prior, *supra* note 124, at 107.

[171] It is what people want.

[172] Barker, *supra* note 169.

[173] *Id.*

[174] Kolbert, *supra* note 94.

[175] *See* Barker, *supra* note 169.

Ratings also determine the messages that are amplified.[176] The chase for ratings creates an argument culture of "debate shows" and blockbuster stories, and the needed element of a public forum to address the important concerns facing the nation continues to be missing.[177] The news media thrives on conflict precisely because conflict attracts viewers, listeners, and readers to the media; the more intense the conflict, the greater the audience; the greater the audience, the higher the ratings, and high ratings represent enormous financial success for media companies and their advertisers.[178] Moreover, the rise of digital technologies and news aggregators is turning everyone into a media company, which fosters more extreme competition to gain followers.[179]

This extreme commercialism proliferates dangerous politics.[180] To illustrate, CNN and other stations obsessively covered Flight 370 because of Americans' addiction to conflict-style reporting.[181] In the meantime, serious global and domestic issues were poorly reported, including the Russia-China oil trade deal that threatened American's petrodollars in the Ukraine crisis.[182] As Kovach stressed, a journalist's *first* loyalty must be to the citizens because poor coverage by journalists produces an unformed citizenry.[183]

B. *Democracy in Danger*

With growing polarization and mistrust, problems begin to seem unsolvable and compromise is not presented as a legitimate option.[184] The commercial media's profit motives must be decoupled from the news because it harms America's fundamental democracy.[185] The policy intervention and structural overhaul needed, like creating safeguards for responsible and informative media, is not within the scope of the article. Sensationalist and personalized news only perpetuate polarization in America.[186] The polarized debate online and on television screens disenfranchises people from authentic public discussion, and cynicism corrodes the quality of civil discourse in the country, threatening the foundation of democratic institutions.[187]

---

[176] Jennifer Akin, *Mass Media*, BEYOND INTRACTABILITY (Mar. 2005), http://www.beyondintract ability.org/essay/mass-communication.

[177] *See generally* KOVACH & ROSENSTIEL, *supra* note 5.

[178] Akin, *supra* note 176.

[179] Anand, *supra* note 121; *see also* Bixby, *supra* note 146.

[180] Victor Pickard, *The Problem with Our Media Is Extreme Commercialism*, THE NATION (Jan. 30, 2017), https://www.thenation.com/article/the-problem-with-our-media-is-extreme-commercialism.

[181] Roberts, *supra* note 166.

[182] *Id.*

[183] *See* KOVACH & ROSENSTIEL, *supra* note 5, at 5.

[184] *Id.* at 174–75.

[185] Pickard, *supra* note 180.

[186] *See* Anand, *supra* note 121.

[187] Robert Berdahl, Speech Delivered to American Society of Newspaper Editors Credibility Think Tank (Oct. 15, 2004).

146   *Pepperdine Dispute Resolution Law Journal*   [Vol. 19:1

Sensationalism in news media prevents the public from being knowledgeable participants in policy conversations, and a democracy depends on informed citizens that deliberate and determine the best policy solutions.[188] While electoral politics naturally creates conflict with embedded "winners" and "losers" of a political race, public life and government do not have to be conflict-ridden.[189] Legislation may entail "maneuvers and showdowns but the most important steps often come when people find areas where they can agree."[190] Progress can be made.

As the fourth branch of government, media professionals have an opportunity to mediate progress by carefully choosing *what* they report and *how* they perform and conduct their reporting duties. America needs honest, factual stories, and reporters who can give us the news in a responsible and trustworthy manner. ADR techniques used by mediators, in the next section, provide an understanding about the psychology of conflicts and recommendations to manage conflict to foster an understanding of perspectives in our polarized climate. In the words of Seth Godin:

Giving the people what they want isn't nearly as powerful as teaching people what they need. There's always a shortcut available, a way to be a little more ironic, cheaper, more instantly understandable. There's the chance to play into our desire to be entertained and distracted regardless of the cost. Most of all, there's the temptation to encourage people to be selfish, afraid, and angry. Or you can dig in, take your time, and invest in a process that helps people see what they truly need. When we change our culture in this direction, we're doing work that's worth sharing. But it's slow-going. If it were easy, it would have happened already. It's easy to start a riot, difficult to create a story that keeps people from rioting. Don't say, 'I wish people wanted this.' Sure, it's great if the market already wants what you make. Instead, imagine what would happen if you could teach them why they should.[191]

VI.   RECOMMENDATIONS ON MEDIA'S USE OF ADR TO ENHANCE UNDERSTANDING

A.   *Media as Mediators*

Today, political dialogue is full of emotion instead of respectful debate, sensationalism instead of honest reporting. The media can use its power to increase polarization and extremism by marginalizing certain groups and only

---

[188] David F. Ransohoff & Richard M. Ransohoff, *Sensationalism in the Media: When Scientists and Journalists May Be Complicit Collaborators,* 4 EFFECTIVE CLINICAL PRAC. 185, 185 (2001).

[189] FALLOWS, *supra* note 163, at 163.

[190] *Id.* at 163.

[191] *Cartographer of Meaning in a Digital Age,* ON BEING (Jan. 5, 2017), https://onbeing.org/programs/maria-popova-cartographer-of-meaning-in-a-digital-age-jan2017/ (quoting Seth Godin, *Give the People What They Want,* SETH'S BLOG (Feb. 4, 2015), https://seths.blog/2015/02/give-the-people-what-they-want/).

quoting extreme members and positions.[192] Or the media can be a catalyst for a change.

With polarization in America at its zenith, media professionals have a unique opportunity to use their role as news people, daresay as mediators,[193] to create knowledgeable participants in policy conversations with an understanding of diverse perspectives. As the fourth branch of government, the media's role is to check on the abuses of power and be an educator of the people.[194]

To educate a polarized public, journalists and news reporters should develop skills to be better mediators and adapt to the challenges facing the twenty-first century. Principles of sound mediation mirror principles of sound journalism.[195] A mediator facilitating dispute resolution between parties and a journalist in the public eye uncovering the truth are alike in many ways and different in other ways.[196] Meditation is an intense process "where parties, counsel, and the mediator often spend hours locked in conference rooms attempting to hash out the details of a proposed settlement."[197] In this process, the mediator becomes aware of private details of each party's position.[198] Similarly, in the process of news gathering, a journalist becomes aware of the detailed accounts of each side and must determine how to frame the information into the news the public consumes.[199]

Like a mediator, journalists listen to both, if not multiple, sides of a conflict while maintaining detachment and neutrality.[200] Like a journalist, a mediator brings parties together, gains information, and provides realistic evaluation while refereeing the process.[201] On the other hand, mediation is different from journalism in that mediations are held in private with the parties' consensus to work towards a settlement, whereas a media professional's work is generally in the public purview with the goal of uncovering truth (although confidential interviews or informant conversations may be in private) with no such agreement.[202] Perhaps, movement towards a resolution and "truth-finding" for the news are sometimes incompatible objectives; nevertheless, meditation techniques can help uncover citizens'

---

[192] Melissa Baumann & Hannes Siebert, *The Media as Mediator*, NIDR FORUM 28, 28 (Winter 1993).

[193] The media unavoidably, necessarily mediates conflicts. Baumann & Siebert, *supra* 192, at 28.

[194] Coleman, *supra* note 15, at 247, 252.

[195] Baumann & Siebert, *supra* note 192, at 28.

[196] Pauli, *News Media as Mediators*, *supra* note 1, at 719.

[197] LEAH M. QUADRINO, COMPLEX US MEDIATION: KEY ISSUES AND CONSIDERATIONS, Practical Law Practice Note 1-575-6667 (2014).

[198] *Id.*

[199] *See* Shoemaker *supra* note 4, at 109.

[200] Pauli, *News Media as Mediators*, *supra* note 1, at 719.

[201] *Id.* at 720.

[202] *Id.* at 721–22; KOVACH & ROSENSTIEL, *supra* note 5, at 5–6, 167.

148   *Pepperdine Dispute Resolution Law Journal*   [Vol. 19:1

underlying interests to facilitate accurate and honest reporting and develop understanding for all parties involved.[203]

Techniques of ADR professionals who facilitate high-conflict mediations and negotiations can equip journalists in the impossible task of presenting the truth and educating a polarized public. There are a litany of training manuals, articles, and books on the principles of ADR and on the art of negotiation and mediation. This article will focus on only three areas: (1) a mediator's presence and growing self-awareness, (2) understanding conflict and cognitive biases, and (3) managing an impasse.

B. *A Mediator's Presence & Growing Self-Awareness*

"Real experts ... are intellectually honest and brutally self-critical with themselves. They examine their mistakes squarely, deconstruct them, and relentlessly search for the impeccable."
– Peter Adler [204]

As many experienced mediators will agree, mediation is more than a "bag of tricks" or techniques.[205] The interplay of one's psychological, intellectual, and spiritual qualities has a direct impact on mediation, and by extension, the effectiveness of news reporting.[206] As mediators Daniel Bowling and David Hoffman assert, "this impact may be one of the most potent sources of the effectiveness of mediation."[207]

Career growth for a mediator (or newsperson) is a process of evolution. At the start, the mediator's goal is to study and practice techniques; next, the mediator tries to gain a deeper understanding of how mediation (or the news) works; lastly, the mediator strives to deepen an "awareness of how his or her personal qualities—for better or worse—influence the mediation [or news reporting] process."[208] A "mediator presence," as experienced by the parties, communicates a message.[209] Similarly, news personalities exert personal influence by their mere presence in disputes, such as in a dispute between party pundits or representatives.[210] When media professionals are at peace with themselves and the world, they subtly carry that peace into the room,

---

[203] *See generally* Pauli, *Transforming News, supra* note 9.

[204] Peter Adler, Chapter 2 in Daniel Bowling & David Hoffman, *Bringing Peace into the Room: The Personal Qualities of the Mediator* (2003), http://www.eyeofthestormleadership.com/pg16.cfm.

[205] LAURENCE J. BOULLE, MICHAEL T COLATRELLA JR., & ANTHONY P. PICCHIONI, MEDIATION: SKILLS AND TECHNIQUES 11 (2008).

[206] *See* DWIGHT GOLANN & JAY FOLBERG, MEDIATION: THE ROLES OF ADVOCATE AND NEUTRAL 98 (Wolters Kluwer eds., 3d ed. 2016) (quoting Daniel Bowling & David Hoffman's article *Bringing Peace into the Room*).

[207] *Id.* at 99.

[208] *Id.*

[209] *Id.* at 99–100.

[210] *Id.* at 100.

interview, or live broadcast and can orient parties in a positive direction.[211] Self-awareness is key.[212]

What does this mean for a media professional reporting the news? Principally, it requires considering and managing personal biases—the lenses through which one sees the world—because bias impacts objectivity in news narratives and impairs the ability to understand the parties' perspectives.[213] Second, it demands a strong emotional intelligence, or EQ, because knowing one's emotions and managing them, as well as recognizing strong emotions and managing relationships with others, is inseparable from balanced reporting.[214]

Strong emotions can sometimes lead to ill-advised conclusions that are counterproductive or even harmful, such as (1) a distorted view, (2) validation of only supporting evidence, or (3) negative reactions to suggestions or evaluations from the other side.[215] A media professional's personal reactions to an issue and ability to manage the responses of others in a news context unequivocally sends a message to the journalist's audience. This is the core of the "mediator's presence" or "newsperson's presence."

Bowling and Hoffman explain that "[u]ntil we develop emotional self-awareness, we will project our own unrecognized emotions onto others."[216] Knowing oneself is to understand one's impulses in a heated debate or disagreeable interview: whether to control the process, react against the people, or rescue the underdog.[217] It is to be aware of identity: views of self, values, culture, and attitude.[218] It is to know if one holds strong convictions against an idea, experience or people, and guard against an unwarranted response.[219]

Lastly, self-reflection, before, during, and after the process of news reporting will be critical in the effort to curb the individual biases that taint objectivity and balance.[220] Consistently evaluating one's "practice" in gathering and reporting the news will develop greater EQ competency.[221]

---

[211] *Id.* at 98, 100.

[212] *See* CRAIG E. RUNDE & TIM A. FLANAGAN, BECOMING A CONFLICT COMPETENT LEADER 4 (2007).

[213] *See* GOLANN & FOLBERG, *supra* note 206, at 99–100.

[214] *See* Louise Phipps Senft, *The Interrelationship of Ethics, Emotional Intelligence and Self Awareness*, ACRESOLUTION (Spring 2004), https://www.baltimoremediation.com/articles/ACR_Ethics.pdf.

[215] *See* Yona Shamir, UNESCO-IHP, *Alternative Dispute Resolution Approaches and their Application in Water Management: A Focus on Negotiation, Mediation, and Consensus Building*, http://www.un.org/waterforlifedecade/water_cooperation_2013/pdf/adr_background_paper.

[216] *See* GOLANN & FOLBERG, *supra* note 206, at 99.

[217] *See* Mieke Brandon, *Refection and Self Awareness*, Address at 11th National Mediation Conference in Sydney 1 (2012).

[218] *See id.* at 1.

[219] *See id.* at 4.

[220] *See id.* at 1.

[221] *See id.* at 5.

News professionals also need competency on the conflict and cognitive biases that are at play with themselves, pundits, interviewees, and the general consumers of the news.

C. *Understanding Conflict & Cognitive Biases*

Conflict is an inescapable part of life. It can be defined as "any situation in which people have incompatible interests, goals, principles, or feelings."[222] A cursory glance at the news will accentuate the landmines of conflict entrenched in our political discourse. Unintentionally, people fall prey to the "drama triangle" of victim, villain, and hero.[223] People in a conflict see the other side as the villain and erect walls of judgment that prevent parties from moving past their "positioning" to consider the possibility of a resolution.[224] This understanding of conflict supports the level of discontent and personalization of politics, including the mutually perceived incompatibility, between the Democrats and the Republicans in America today.[225]

Another way to conceptualize conflict is in three dimensions: behavioral, emotional, and cognitive.[226] The behavioral element refers to the concrete elements surrounding a conflict, such as the Dakota Access Pipeline and Standing Rock Sioux tribe.[227] The emotional element refers to the feelings associated with the incident(s) at any stage of the process, such as a legitimate fear that grows into anger or resentment.[228] And lastly, the cognitive dimension involves how people contemplate, understand, and interpret the issues.[229] Different parties have differing interpretations, and as a reporter it is important to understand and share these varying interpretations and the drivers behind them.[230]

Cognitive biases are "universal human tendencies to process information in ways that often lead to erroneous judgments of others."[231] These judgments arise from incorrect assumptions concerning motivations when ignorant of a person's true intentions.[232] Cognitive biases generally serve as instigators of

---

[222] RUNDE & FLANAGAN, *supra* note 212, at 4.

[223] GARY HARPER, THE JOY OF CONFLICT: TRANSFORMING VICTIMS, VILLAINS AND HEROES IN THE WORKPLACE AND AT HOME 8 (2004).

[224] *See id.* at 8.

[225] *See generally* Pew, *Polarization in America, supra* note 1; Pew, *Divide on Political Values Grows Wider, supra* note 53; Pew, *On the Eve of Inauguration, supra* note 41.

[226] Leonard L. Riskin, *Eleven Big Ideas About Conflict: A Superficial Guide for the Thoughtful Journalist*, J. DISP. RESOL. 157, 159 (2007), http://scholarship.law.ufl.edu/facultypub/456.

[227] *See id.* at 159.

[228] *See id.* at 159.

[229] *See id.* at 159.

[230] *See id.* at 159.

[231] VICTORIA PYNOCHON & JOE KRAYNAK, SUCCESS AS A MEDIATOR FOR DUMMIES 151 (2012).

[232] *Id.*

disputes and often cause negotiations to fall apart.[233] Thus, a newsperson should be cognizant of cognitive biases that cause people in the same situation to view the situation vastly differently. Common cognitive biases include: selective perception, confirmation bias, attribution bias, anchoring, and reactive devaluation.[234]

1. *Selective Perception*

Selective perception underpins why people view the same event differently.[235] When people process new information, they automatically view it from a particular lens and judge the situation from that frame.[236] Data that is inconsistent with the frame is likely to be disregarded.[237] It is the root of why people in arguments can instinctively tune into concepts that support their viewpoint and point out weakness in the other side, but miss weaknesses in their own argument.[238] Understanding selective perception can prompt a reporter to account for bias by asking probing questions to their interviewees.[239]

2. *Confirmation Bias*

Confirmation bias, as discussed earlier, is similar to selective perception.[240] It is the tendency to weigh supporting evidence more heavily and discredit conflicting evidence.[241] Confirmation bias is perpetuated by advocacy or "positioning."[242] For example, positioning occurs when a political party representative takes a "position" while speaking to a reporter or live audience, and the representative focuses on building her case rather than evaluating it fairly because of her commitment to her argument's superiority.[243]

3. *Anchoring*

Anchoring occurs when people form value estimations or benchmarks based on their individual experiences and understanding.[244] Once the initial

---

[233] *Id.*

[234] *See* GOLANN & FOLBERG, *supra* note 206, at 55–60.

[235] *Id.* at 57.

[236] *Id.*

[237] *Id.*

[238] *Id.*

[239] *Id.* at 57.

[240] *Id.*

[241] *Id.*

[242] *Id.*

[243] *See id.*

[244] *Id.* at 58.

benchmark is made, it is difficult to readjust; people are "anchored" by the baseline, even if the comparison or starting point was unsubstantiated.[245] In news coverage, anchoring is evident when political parties are "posturing." For example, Republicans assert that Obamacare was a failure that ruined healthcare and the economy, while Democrats argue it saves lives and helps those who cannot afford insurance.[246] Whether there is a legitimate basis for the numbers or the "positioning," the parties are stuck or "anchored" in their initial positioning.

4. *Attribution Bias*

Attribution bias is the tendency to assume the worst about the opposition.[247] Any action or conduct by the other side is judged in the worst possible light—intentional malfeasance—whereas any ambiguous conduct by one's own side is mere mistake or unintentional conduct.[248] In the political arena, this is apparent. The Democratic Party will scourge the Republican Party for being anti-women's rights but at the same time not praise strong women leaders who identify as Republicans.[249] Similarly, the Republican Party will condemn the Democratic Party over their government "handouts" or "abortions" but often promote wheat subsidies for many of the same philosophical reasons the Democrats push welfare programs.[250]

5. *Reactive Devaluation*

Lastly, reactive devaluation is another level of attribution bias.[251] Any offer or suggestion made by the opposing side is viewed automatically as insufficient or ill-intentioned.[252] Even if, from a neutral perspective, a decent compromise or proposal is offered, reactive devaluation is the tendency to assume the other side does not have one's best interests at heart; therefore, there must be an underlying trap or hidden agenda.[253] News people, like mediators, have a unique role and opportunity to evaluate the merits of

---

[245] *Id.*

[246] *See e.g.,* Erika Franklin Fowler et al., *Media Messages and Perceptions of the Affordable Care Act During the Early Phase of Implementation,* J. HEALTH POL. POL'Y L. 167, 170 (2017).

[247] GOLANN & FOLBERG, *supra* note 206, at 60.

[248] *Id.*

[249] *See, e.g.,* S.E. Cupp, *No, Not All Women are Democrats,* N.Y. TIMES (June 16, 2018), https://www.nytimes.com/2018/06/16/opinion/sunday/no-not-all-women-are-democrats.html.

[250] *See, e.g.,* Erika Eichelberger, *The Four Most Hypocritical Provisions in the GOP Farm Bill,* MOTHER JONES (Nov. 7, 2013), https://www.motherjones.com/food/2013/11/republican-farm-bill-food-stamps-subsidy-hypocrisy/.

[251] *See* GOLANN & FOLBERG, *supra* note 206, at 60–61.

[252] *See id.*

[253] *See id.*

proposals or ideas in the abstract, and being unattached from party affiliation allows them to critically examine a matter and avoid reactive devaluation.[254]

D. *Managing an Impasse*

Awareness of cognitive biases is one aspect, but mediators also have techniques to manage an impasse that arises, which can be useful to the media in news reporting. Key techniques are as follows: emphatic listening, reframing, and reality testing.[255]

1. *Empathetic Listening*

"The most basic of human needs is the need to understand and be understood. The best way to understand people is to listen to them."[256] Mediators and journalists alike actively seek to learn people's stories.[257] By creating the space to be fully heard, a mediator and a journalist can redirect confrontational energy to foster collaboration and openness.[258] Empathy is the ability to project oneself into the other person's shoes to understand their emotions and feelings.[259] Empathetic listening uses verbal and nonverbal cues such as saying, "I see" or nodding to reassure the speaker that the speaker can share without fear of criticism, judgment, or interruption.[260] Active listening requires stopping one's inner conversation or urgings to prepare a response, and truly listening to the other person.[261]

An extension of the technique would be reflective listening—where the listener summarizes back what he or she heard to demonstrate understanding.[262] It is important to note, empathy does not equal sympathy.[263] Empathy does not mean that one agrees with the speaker; it simply means that the listener indicates that he or she has heard and understood the speaker.[264] Entertainment news today unfortunately uses mostly confrontational conversations playing a "game of ping-pong marked by a series of 'yes, but' as the person immediately rebuts the other's points."[265] While perhaps

---

[254] *See id.* at 160–61.

[255] *Id.* at 123–32.

[256] Dr. Ralph Nichols, INT'L LISTENING ASS'N, https://www.listen.org (last visited Feb. 12, 2018).

[257] HARPER, *supra* note 223, at 93.

[258] *See id.*

[259] GOLANN & FOLBERG, *supra* note 206, at 125.

[260] *Id.* at 125–26.

[261] *See* HARPER, *supra* note 223, at 93.

[262] GOLANN & FOLBERG, *supra* note 206, at 126.

[263] *Id.* at 126.

[264] *Id.* at 126.

[265] *See* HARPER, *supra* note 223, at 95.

154   *Pepperdine Dispute Resolution Law Journal*   [Vol. 19:1

entertaining, this style can be damaging and fails to create an understanding of perspectives.[266]

On the other hand, shows like *Messy Truth* with Van Jones or *On Being* with Krista Tippet demonstrate unique, non-partisan styles of reporting aimed at building understanding through conversations with real citizens.[267] Each reporter has their own style, but their attention to listening goes a long way in building understanding.[268]

2. *Reframing*

Mediators manage conversations by reframing, in other words restating or paraphrasing, disagreement using nonjudgmental language.[269] Reframing can also be defined as using different words, concepts, or emphases to appropriate for the context.[270] Journalists aim to gather information to report to the public, not understate disagreement.[271] However, this does not mean reframing is useless. Especially when parties are in a joint conversation with the journalist, the journalist may find that using the reframing technique to rephrase statements of one side's "position" in terms of "interests" can be powerful:[272] for example, framing "immigrants" as people seeking a better life or escaping poverty, or "gun-lovers" as individuals wanting the right to keep their family safe or continue a hard-earned way of life without interference. Reframing can help the journalist and the participants uncover deeper interests and foster understanding even if they do not agree.[273] It may even create space for a journalist to probe one's side about their understanding of the other side's positions, and even potentially create workable solutions.[274]

3. *Reality Testing*

Reality testing is used by mediators to guide parties to uncover gaps in their understanding of a situation.[275] In serving as an agent of reality, the mediator (and the journalist) can pose critical questions and help parties

---

[266] GOLANN & FOLBERG, *supra* note 206, at 126.

[267] *See* Van Jones, *Beyond the Messy Truth*, http://www.vanjones.net/beyond_the_messy_truth (Jan. 28, 2018); Krista Tippet, *On Being, Civil Conversations Project*, http://www.civilconversationsproject.org (Feb. 8, 2018).

[268] GOLANN & FOLBERG, *supra* note 206, at 126 (noting that good listening helps bring understanding between parties).

[269] Pauli, *News Media as Mediators, supra* note 1, at 725–26.

[270] *See* BOULLE, *supra* note 206, at 128.

[271] Pauli, *News Media as Mediators, supra* note 1, at 725–26.

[272] *Id.* at 725–26.

[273] *See* GOLANN & FOLBERG, *supra* note 206, at 130–31.

[274] *See* Pauli, *News Media as Mediators, supra* note 1, at 726 (arguing against conflict oriented frames because collaborative spaces foster better understanding and actual resolution of issues).

[275] *See* BOULLE, *supra* note 206, at 233.

conduct evaluations of the issue.[276] In helping a party see the merits of an issue objectively, the party can move towards a more realistic view of the circumstances as opposed to having a view that is completely distorted by cognitive biases.[277] The newsperson, using reality testing, can ask a party to respond to the opponent's claims, illustrating the strengths and weaknesses in the party's own position.[278] In a broadcast or written piece, this critical engagement would be illuminating to the public consuming the news. However, it is important to note that this technique is more interventionist and commonly conducted in private caucusing in mediation to avoid creating vulnerability and defensiveness in the party who is being challenged.[279] As a reporter, critical evaluation jointly with a party can foster understanding, but there is also risk the silent audience may not also join the ride; but at the very least, the journalist has fulfilled her role in enhancing the public understanding of viewpoints.

VII. CONCLUSION

The intended role of the media, as the unofficial fourth branch of the government, is to arm the public with the information they need to be free and self-governing.[280] Unfortunately, the rise of technology, confirmation bias, and mistrust of the media has led to an increasingly polarized America. The growing personalization of politics and misunderstandings demands that the media bolster their toolkit with ADR insights and methods in order to promote understanding. This article provides a few techniques that may prove useful. Every news media professional must grasp that "public discourse lies at the heart of and actually predates formal American journalism."[281] We need more understanding across the aisle today and the Media's adoption of ADR techniques may be the best first step towards a better informed, more understanding public.

---

[276] *See id.* at 234.

[277] *See id.*

[278] *See id.* at 233.

[279] *Id.* at 235.

[280] *The 4th Estate as the 4th Branch,* TEACHING HISTORY, https://teachinghistory.org/history-content/ask-a-historian/23821 (last visited Sep. 27, 2018).

[281] KOVACH & ROSENSTIEL, *supra* note 5, 170.

Why We Can't "Just All Get Along"[1]: Dysfunction in the Polity and Conflict Resolution and What We Might Do About It

Carrie Menkel-Meadow[2]

I. THE TROUBLE WE'RE IN

These are very troubled times. The polity is seriously divided; people who march for white supremacy and hate are called "nice and very good people" by an unhinged, but Constitutionally elected, President;[3] relations between citizens of color and police are at a high level of hostility and distrust; Congress is unable to pass virtually any legislation; and policy differences over immigration, trade, taxation, and health care are so great that even a ruling party cannot get anything done. Ordinary citizens can learn the news and follow their political preferences by insulating themselves in media "bubbles," seeking out websites, cable news channels and social media outlets, and even physical neighborhoods, that are consistent with their own views.[4] Universities, presumed sites of "free speech" and intellectual engagement are now marked by limited conversations mired in claims of "political correctness" and "trigger warnings,"[5] all of which have sharply limited our collective ability to talk to and really listen to each other. We, as a nation, (and perhaps

---

1. This is a reference to the plaintiff cry of Rodney King, a black man, beaten by police in Los Angles on March 3, 1991, resulting in days of civil unrest in Los Angeles and leading to two different prosecutions and lawsuits against the culpable police officers (in one of the first police abuse incidents to be fully captured on videotape). Karen Grigsby Bates, *When LA Erupted in Anger: A Look Back at the Rodney King Riots*, NPR (April 26, 2017, 1:21 PM), http://www.npr.org/2017/04/26/524744989/when-la-erupted-in-anger-a-look-back-at-the-rodney-king-riots. (I was hosting my Negotiation and Conflict Resolution UCLA law students for dinner at my home in Los Angeles when rioting broke out after the acquittal verdict of the police officers charged with the beating and the city was on curfew for some days after that. "Why can't we all just get along" became a popular call for both citizens and professionals in dispute resolution to seek ways to encourage productive dialogue across political and social differences.

2. Chancellor's Professor of Law and Political Science, University of California, Irvine and A.B. Jr. Professor of Law, Dispute Resolution and Civil Procedure, Emerita, Georgetown University Law Center. I thank my research assistant and conflict resolver in chief, and fellow traveler in all of life, Robert Meadow, for critical (and also supportive) comments and suggestions.

3. President Donald J. Trump, Press Conference at Trump Tower, New York (Aug. 15, 2017)(remarks made following "alt-right" march and protest in Charlottesville, Virginia in which three people died and nineteen were injured in a march to protest the taking down of Confederate General Robert E. Lee statue on August 12, 2017).

4. *See* Andrei Boutyline & Robb Willer, *The Social Structure of Political Echo Chambers: Variations in Ideological Homophily in Online Networks*, 38 POL. PSYCHOL. 551, 554 (2017); *See also* Shanto Iyengar & Sean Westwood, *Fear and Loathing Across Party Lines: New Evidence on Group Polarization*, 59 AM. J. POL. SCI. 690, 691 (2015); BILL BISHOP, THE BIG SORT: WHY CLUSTERING OF LIKE-MINDED AMERICAN IS TEARING US APART 6 *(2008)*.

5. *See generally* MARK HUME, TRIGGER WARNINGS: IS THE FEAR OF BEING OFFENSIVE KILLING FREE SPEECH? (2015); *see also* Jeannie Suk, *Shutting Down Conversations About Rape at Harvard Law School*, THE NEW YORKER (December 11, 2015), https://www.newyorker.com/news/news-desk/argument-sexual-assault-race-harvard-law-school; ERWIN CHEMERINSKY & HOWARD GILLMAN, FREE

even more fraught, as it grows increasingly dangerous as can be seen recently in more violent confrontations at demonstrations around the nation on Confederate statue removal,[11] racialized policing, immigration, women's rights and "free speech,"[12] not to mention how increased bellicosity at the international level (US-North Korea, Russia-Ukraine, Syria) have put us on a military (and possibly nuclear)[13] precipice not seen since the Cold War. I have written about the dangers of brittle adversarial thinking and behavior in the legal system and polity for decades,[14] now seemingly to little avail. Whatever we, problem solving negotiators, mediators, deliberative democrats, and public policy facilitators have accomplished in particular cases and matters,[15] our contributions to major culture change and paradigm shifts to thinking about each other and the world with a different framework, have largely failed. Here I offer some explanations of why, and what we might still do about it, though I remain sadly skeptical of my own optimism. We must bring together the learning and data of social science and behavioral research on how humans behave and process information together with the particular tools and techniques (and theories) of conflict resolution professionals. Either alone is not enough.

II. OUR DIFFERENCES CANNOT JUST BE "REASONED" AWAY

My basic point is simple: *rationality* will not necessarily bring us together. While much of legal thought, science, philosophy and Enlightenment-era epistemology has promised us that *reasoned argument,* when properly structured for larger deliberative settings in the polity,[16] will persuade decision makers (whether

11. *See, e.g.,* Lulu Garcia Navarro, *Richmond, Va., Grapples With Fate of Confederate Monuments,* NPR (Aug. 13, 2017, 8:33 ET), http://www.npr.org/2017/08/13/543197271/richmond-va-grapples-with-fate-of-confederate-monuments; *The Alt-Right on Campus: What Students Need to Know,* SOUTHERN POVERTY LAW CENTER, (Aug. 10, 2017), https://www.splcenter.org/20170810/alt-right-campus-what-students-need-know.

12. A new issue is emerging to claim that regulation of some marches and demonstrations may be constitutionally permissible, e.g. the banning of weapons at mass gatherings, as the First Amendment protects only "peaceable assemblies," though any such regulation is likely to lead to litigation on the potential clashes of First and Second Amendment interpretations. *See* L.A. Times Editorial Board, *Editorial: Don't Restrict Free Speech. Restrict the Right to Carry Guns at Potentially Explosive Public Events,* L.A. TIMES (Aug. 23, 2017), http://www.latimes.com/opinion/editorials/la-ed-firearms-charlottesville-texas-protest-20170823-story.html.

13. Choe Sang-Hun & David E. Sanger, North Korea Fires Missile Over Japan, N.Y. TIMES (Aug. 28, 2017), https://www.nytimes.com/2017/08/28/world/asia/north-korea-missile.html.

14. *See generally* Carrie Menkel-Meadow, *The Lawyer's Role in Deliberative Democracy,* 5 NEVADA L. J. 347 (2004-2005); *see also* Carrie Menkel-Meadow, *The Trouble with the Adversary System in a Post-Modern, Multi-cultural World,* 38 WM. & MARY L. REV. 5 (1996); Carrie Menkel-Meadow, *Deliberative Democracy and Conflict Resolution: Two Theories and Practices of Participation in the Polity,* ABA DISP. RESOL. MAG. (2006), at 18.

15. For some great success stories, see SUSAN L. PODZIBA, CIVIC FUSION: MEDIATING POLARIZED PUBLIC DISPUTES (2012); *see also* LAWRENCE SUSSKIND, SARAH MCKEARNAN & JENNIFER THOMAS-LARMER, THE CONSENSUS BUILDING HANDBOOK: A COMPREHENSIVE GUIDE TO REACHING AGREEMENT [hereinafter CONSENSUS BUILDING HANDBOOK] (1999); CAROLYN J. LUKENSMEYER, BRINGING CITIZEN VOICES TO THE TABLE (2013); JOHN FORESTER, DEALING WITH DIFFERENCES: DRAMAS OF MEDIATING PUBLIC DISPUTES (2009).

16. *See* JÜRGEN HABERMAS, BETWEEN FACTS AND NORMS: CONTRIBUTIONS TO A DISCOURSE THEORY OF LAW AND DEMOCRACY (1996); *see also* DELIBERATIVE DEMOCRACY (James Bohman & William Rehgs eds., *1997);* JOHN DRYZEK, DELIBERATIVE DEMOCRACY AND BEYOND: LIBERALS, CRITICS, CONSTESTATIONS (2000); DELIBERATIVE DEMOCRACY (Jon Elster ed., 1998). AMY GUTMANN & DENNIS THOMPSON, WHY DELIBERATIVE DEMOCRACY? (2004); DELIBERATIVE DEMOCRACY

legislators, judges, or citizen-voters) to use facts and persuasion to reach good, if not perfect, decisions, the truth is that human beings actually use both more and "less" than reason in their thinking and decision making processes. Like those who raise their voices in English when traveling in other countries, expecting that louder English will make those who don't speak it understand them better, much of deliberative democracy, and attempts to facilitate difficult political conversations, assumes that "more reason" and more and better arguments (and facts) will bring "others" around to see "the truth," the "light," or the correct or best policy.

For example, in a recent article in the *Los Angeles Times*, advice given on "how to talk to climate change skeptics"[17] suggested using seven arguments, based on science (take the long data view; don't let skeptics cherry pick counterexamples; see the effects of greenhouse gases in particular, produced increasingly, since coal use began, by human agency; note the decreased solar activity; demonstrate the depth and extent of recent changes, beyond the "normal" variations; see that even if the Antarctic is not diminishing (it is, actually, I was just there), the Arctic is; and see that even though polls show that most people don't know it, more than 95% of all climate scientists agree on its causes and effects and dangers). Although the article offered some good advice, "always ask skeptics about the data their argument is based on" (see below), it concluded with "as with any contentious issue, you have to realize when you are talking to someone who just wants to argue, and you should just stop." In other words, if your rational arguments and science are not persuasive, just stop. This is better than religious warfare of most of human history when people killed those with whom they didn't agree,[18] but even if you are from the "science" camp (as I am), is this the only way to get people to "understand" or rethink what they think or believe? Recently many commentators pointed out that many people "selectively believed" in science to expect and observe the total eclipse as predicted, and also did not doubt the arrival of Hurricane Harvey (although debates remain about whether this hurricane was "worse" because of climate change [more warm moisture in the air] or more damage resulted because of more development on flat land). My own view, as often, is that both can be true at the same time. So "reasoning" alone will not necessarily help people get along.

Instead, modern intellectual and scientific challenges to pure rationality include the work of cognitive scientists in chronicling "deviations" from rational thinking in common human cognitive errors[19] (including such now familiar "errors" as biases

---

HANDBOOK: STRATEGIES FOR EFFECTIVE CIVIC ENGAGEMENT IN THE TWENTY-FIRST CENTURY (John Gastil & Peter Levine, eds., 2005); DAVID KAHANE, DANIEL WEINSTOCK, DOMINQUE LEYDET & MELISSA WILLIAMS, DELIBERATIVE DEMOCRACY IN PRACTICE (2010). For an argument that deliberative democracy theorists are somewhat incompatible with conflict resolution theorists and practitioners, see Hiro N. Aragaki, *Deliberative Democracy as Dispute Resolution? Conflict, Interests and Reasons*, 24 OHIO. ST. J. DISP. RES. 407 (2009) (arguing that deliberative democrats regard interest-based "bargaining" as inferior to the public-minded searches for the "common good" through deliberation through reasoned argument).

17. LA Times, How to talk to Climate Change Skeptics, September 2, 2017, California Section, Science File, page B-2.

18. Still going on, of course, in some parts of the world.

19. *See generally* DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2013); *see also* RICHARD NISBETT & LEE ROSS, HUMAN INFERENCE: STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT (1980); JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982); PAUL BREST & LINDA HAMILTON KRIEGER, PROBLEM SOLVING, DECISION MAKING AND PROFESSIONAL JUDGMENT (2010).

from vividness, primacy, recency, endowment effects, statistical probability illiteracy, reactive devaluation, anchoring and many others), post-modern questioning of "absolute" truths in theories of hermeneutic interpretations,[20] and most relevant to this essay, the power of emotions,[21] including religious commitments, feelings (fear and threat,[22] or generosity, and shared empathy,[23] among others) and *affective adherence* to deeply held and widely disparate values, which are often counter to or inconsistent with "rational" truths.[24]   Social psychologists have labeled this the "tribal moral divide" where people act on moral "intuitions" before even seeking confirming or disconfirming facts or information, adhering strongly to belief systems, with possibly genetic or at least familial and religious bases.[25]

We, as conflict resolution professionals need to learn to structure communication around these complex ideas, as much as we focus on legal, social and economic issues, parties' needs and interests,[26] and the more instrumental foci of much conflict resolution facilitation.[27]   There are explanations for why "we can't just all get along," having to do with different world views and value systems and assumptions (from religious, familial, political, and ethical commitments) and from widely disparate life experiences, and reactions to those experiences.[28]   (Consider the old saying, from my New York youth, "a liberal is a conservative who hasn't been mugged yet." Assuming, of course, that the experience of being mugged would lead one to stronger law and order values, as opposed to the "restorative justice"[29] and mercy values held by liberals who stay liberal, even after having been mugged!) Most human beings are a complex mix of using their brains (reason), their heart (values, commitments) and their stomachs (can I really digest ("live with") this) when they speak, argue, make decisions, and act.

As I have suggested in many other fora, there are at least three "modes" of discourse in all decision making—the *rational-principled,* (brain) *interest-based bargaining and trading,* (stomach) and the *affective-emotional-value based* (heart)

---

20. *See generally* HANS GEORG GADAMER, THE HERMENEUTIC TURN (2004); *see also* Carrie Menkel-Meadow, *The Trouble with the Adversary System in A Postmodern, Multicultural World,* 38 *Wm. & Mary L. Rev.* 5 (1996).

21. JON ELSTER, ALCHEMIES OF THE MIND: RATIONALITY AND THE EMOTIONS (1999).

22. *See* ABRAMSKY, supra note 10.

23. *See* Douglas H. Yarn & Gregory Todd Jones, *A Biological Approach to Understanding Resistance to Apology, Forgiveness, and Reconciliation in Group Conflict,* 72 L. & CONTEMP. PROBS. 63, 66 (2009); *see also* Carrie Menkel-Meadow, *Is Altruism Possible in Lawyering?,* 8 GA. ST. L. REV. 385 (1992).

24. JONATHAN HAIDT, THE RIGHTEOUS MIND: WHY GOOD PEOPLE ARE DIVIDED BY POLITICS AND RELIGION (2013); *see also* JONATHAN HAIDT, CAN'T WE ALL DISAGREE MORE CONSTRUCTIVELY (2017).

25. *Id.*

26. *See* the teachings of ROGER FISHER, WILLIAM URY, & BRUCE PATTON in the classic, GETTING TO YES: NEGOTIATING AGREEMENT WITHOUT GIVING IN (3d ed. 2011); *see also* Carrie Menkel-Meadow, *Toward Another View of Legal Negotiation: The Structure of Problem Solving,* 31 UCLA L. REV. 754 (1984).

27. Some of those who work (I do not) in psycho-neurolinguistics in our field are already doing some of this, see Yarn & Jones, *supra* note 23.

28. *Compare* J. D. VANCE, supra note 10, *with* Betsy Rader, *I was born in poverty in Appalachia: Hillbilly Elegy Doesn't Speak for Me,* WASHINGTON POST (Sept. 1, 2017), https://www.washingtonpost.com/opinions/i-grew-up-in-poverty-in-appalachia-jd-vances-hillbilly-elegy-doesnt-speak-for-me/2017/08/30/734abb38-891d-11e7-961d-2f373b3977ee_story.html?hpid=hp_no-name_opinion-card-b%3Ahomepage%2Fstory&utm_term=.ed0ec58303e7.

29. *See generally* Carrie Menkel-Meadow, *Restorative Justice: What Is it and Does It Work?,* 3 *Ann. Rev. L & Soc. Sci.* (2007).

set of claims that people make on each other and within themselves, in different fora:[30]

*MODES OF CONFLICT RESOLUTION**

MODE OF DISCOURSE	PRINCIPLED (REASONS) (Brain)	BARGAINING (INTERESTS) (Stomach)	PASSIONS (NEEDS/ EMOTIONS/ RELIGION) (Heart)
FORMS OF PROCESS:			
Closed	Some court proceedings; arbitration	Negotiation- U.S. Constitution; diplomacy	Mediation (e.g., divorce)
Open	French Constitution; courts; arbitration	Public negotiations; Some labor	Dialogue movement
Plenary	French Constitution	Reg-Neg	Town meetings
Committees	Faculty committees; task groups	U.S. Constitution/ U.S. Congress	Caucuses-interest groups
Expert/Facilitator	Consensus building	Mini-trial	Public conversations
0Naturalistic (Leaderless)			Grassroots organizing/WTO protests; Occupy Wall St.
Permanent	Government, institutions	Business organizations, unions	Religious organizations, Alcoholics Anonymous,
Constitutive	UN, national constitutions	National constitutions/ professional associations	Civil justice movements, peace
Temporary/Ad Hoc	Issue organizations/social justice	Interest groups	Yippies, New Age, vigilantes

Principles = reasons, appeals to universalism, law
Bargaining = interests, preferences, trading, compromises

---

30. *Scaling Up, supra* note 9, at 9-10; *Carrie Menkel-Meadow, From Legal Disputes to Conflict Resolution and Human Problem Solving: Legal Dispute Resolution in a Multidisciplinary Context*, 54 J. LEGAL EDUC. 7, 28 (2004);
CARRIE MENKEL-MEADOW, MEDIATION AND ITS APPLICATIONS FOR GOOD DECISION MAKING AND DISPUTE 37-38 RESOLUTION (2016);
*Partially derived from categories specified by Jon Elster, *Strategic Uses of Argument, in* BARRIERS TO CONFLICT RESOLUTION 236 (Kenneth J. Arrow et al. eds., (1995).

Open = public or transparent meetings or proceedings
Closed = confidential, secret process or outcomes (settlements)

Plenary = full group participation, joint meetings
Committees = task groups, caucuses, parts of the whole

Expert-facilitator = led by expertise (process or substantive or both)
Naturalistic = leaderless, grassroots, ad hoc

Permanent = (Organizational, institutional),
Constitutive = constitutional

Temporary = ad hoc groups or disputants
Some predicted effects of process on outcome:

Closed = (confidential) proceedings allow more expression of interests, needs and passions = more "honest" and candid, allow more "trades," less posturing, open to vulnerability

Open (transparent) proceedings require more principled/reason justifications, produce more rigidity

Until and unless we take all three of these modes seriously to co-exist and be "managed" in political discourse we will not be able to get beyond polarization. We still may not be able to get beyond some basic value polarizations, (ever!), but if we don't learn to deal with these three discourses together, in the same room, in the same conversations, in the same media, we are, I think, doomed to continue to exist in our value-defined echo chambers.

III. CAN WE THINK, FEEL, BELIEVE AND BARGAIN AT THE SAME TIME?

So, the issue for getting along and getting things done is how can we harness conflict resolution theory and practice to provide structures and processes that allow these discourses to be expressed and heard in different places in the polity—from the individual, to the family, to the workplace, to the political party, polling booth, media, legislatures, courts, executive offices, and embassies. For me, the answer is not "blowing in the wind,"[31] but can be found primarily in the empathy and human connections that allow people to talk and listen to each other, across great differences in a mutual desire for some understanding, if not agreement. These days, I find sources and examples of those possibilities more in literature, film, and the arts (rather than in real legal or political practice), where people are actually seen to learn from each other, and, occasionally and hopefully, transcend their committed

---

31. BOB DYLAN, *Blowing in the Wind, on* The Freewheelin' Bob Dylan (Columbia Records1963). The question there is "how long will it take for people to know that too many people have died?" and consider the horrors of war and death and the promise of peace, recounted in several verses and chorus.

understandings of the world. In those rarely glimpsed moments they may come to change their views and appreciate and take seriously "the Other."[32]

When experienced or observed these are like "epiphanies" or "revelatory moments," when people actually learn or radically change their mind or beliefs about something. As you read this, consider, *when did you ever totally change your views about something? How did it happen?*[33] I will review a few of those examples here and suggest how we might harness these examples to conflict resolution theory and practice, for more effective deliberation and decision making in our current polarized climate. And, at the same time, it is important to consider that such "epiphanies" and "revelations" cannot be assumed to be only one way, as is often assumed in our politically "liberal" field.[34] And, such emotionally revelatory moments described as "sacred" or "magical" by many mediators,[35] may also be momentary, if not harnessed to the public sphere from the private (and in political, as well as personal, decision making). So our challenge remains, when and if individualized conflict resolution strategies and tools can work at the individual level of understanding (such as with my peace work in Israel-Palestine at the personal level),[36] how and

---

32. For those readers who are interested, I keep a list of films, novels and other materials which demonstrate such moments, see, for example, DISTURBING THE PEACE (Reconsider, 2016), a recent documentary on the founding of Combatants for Peace (founded by Israelis and Palestinians). As one former Israeli Defense Force solider described, in the moment he was about to deny entrance across a check point for sick Palestinian children, and his wife called with news of his own children, he realized he "couldn't do it anymore." Children are children on both sides.

33. A recent such moment for me came on a return visit to Athens. I spent a full year living next door to the Elgin Marbles in the British Museum in London some years ago and often taught with the formal document used the by the Museum to explain the "dispute" and why the Elgin Marbles were "better off" being taken care of and displayed in London, rather than Athens. Many of my students had written papers and made dispute resolution proposals for a variety of contested art ownership issues (post-colonial, contested provenance, Nazi art looting, University and archeological "thefts" some now governed by treaty. *See* Diane Orson, *Finders Not Keepers: Yale Returns Artifacts to Peru*, NPR (Dec. 18, 2011), https://www.npr.org/2012/01/01/143653050/finders-not-keepers-yale-returns-artifacts-to-peru. It took a visit *in situ* to the new museum across from the Parthenon (where the other "half" of the Marbles are displayed) to change my mind completely. The newly architected building, with archeologists permanently available for conversation, and the "epiphany" of seeing the sublime friezes in their "natural" home, made the "return" of all of them to their historical home seem the best outcome to me. (There are of course, very expensive other outcomes for such disputes- e.g. "rotation" of displays etc., but this time I think I have totally changed my mind). I was moved by the natural and historical habitat and the care with which Greek archeologists and scientists were managing their heritage.

34. In the founding days of ADR it was clear that many of the founders were the products of WWII and the Holocaust, the anti-Nuclear movement (SANE); the civil rights movement, the anti-War (Vietnam War) movement, the feminist movement and now gay and other civil rights movements. In an early meeting of SPIDR (now merged into ACR), we told personal stories and noted similar political commitments or social/familial experiences of alcoholism or abuse that led to a desire to "heal the world" and find "bridges" of understanding, mostly in service to what would be considered liberal or "progressive" values. Someone still needs to do an excavation of the first generation's motivating influences; for just a bit of this, see Carrie Menkel-Meadow, *The Historical Contingencies of Conflict Resolution*, 1 INT'L J. OF CONFLICT ENGAGEMENT AND RESOL. 32 (2013); *see also* Carrie Menkel-Meadow *Why Hasn't The World Gotten To Yes?: An Appreciation and Some Reflections*, 22 NEGOT. J/ 485- (2006); Carrie Menkel-Meadow, *Roots and Inspirations: A Brief History of the Foundations of Dispute Resolution, in* HANDBOOK OF DISPUTE RESOLUTION 13 (Robert Bordone & Michael Moffitt, eds., 2005).

35. *See* Sarah Cobb, *Creating Sacred Space: Toward A Second Generation Dispute Resolution Practice, in Dialogue on the Practice of Law and Spiritual Values* (James F. Henry et al. eds.) 28 FORDHAM URB. L. J. 1017 (2001); see *also* PODZIBA, *supra* note 15.

36. *See* THE PARENTS CIRCLE-FAMILIES FORUM, http://www.theparentscircle.com/ (last visited Nov. 13, 2017); *see also* COMBATANTS FOR PEACE, www.cfpeace.org (last visited Nov. 13, 2017); and films such as ENCOUNTER POINT *(Just Vision Films* 2006); Menkel-Meadow & Irena Nutenko, *The Next Generation: Creating a New Peace Process in the Middle East*, 25(4) NEGOT. JOURNAL 569 (2009).

when can these "moments" be scaled up to the political and policy making level? I offer some ideas here, but I remain skeptical on such deeply divided issues as in our own polity (immigration, health care, and race relations) and other issues on the world stage (EU-Brexit, migration policy, Israel-Palestine, Syria, North Korea, and many other international conflicts).

Since we cannot use the same processes and structures in all settings (individual disputants, families, legal disputants, workers and managers, political parties, communities and nations), we must also be conscious, in our attempts to scale up, of both *"process pluralism,"*—what process is appropriate for what kind of dialogue, conversation, relationship, decision environment or political act—and *"dispute system design"*—how can we as conflict resolution professionals design, build and plan appropriate formats, tools and experiences to encourage constructive and learning settings for real human engagement in such different settings. And we must also recognize that often those moments of "revelation" or deep understanding are more random, individualized moments that we must learn to "capture" and organize in some productive way.

When we see a disagreement, political or relational, what are the sources of the disagreement? Research in political science, sociology, and psychology all consider the impact of groups (sometimes called "tribes") on both ideology (organized frameworks of thought) and individual opinions and actions. In legal and economic situations, the assumptions are most often that some form of self-interest and economic maximization is at play, seemingly assuming that all "rational" actors maximize in more or less the same way and with the same values, assuming conflicts over scarce resources or identity.[37] Having spent years reading, teaching, practicing with, and testing these theories in different contexts, I believe that there are complex interactions among and between *group* or "tribal" (religious, class, ethnicity,) inputs into *individual* thought processes and decision making, and the structures and constraints of *situations,*[38] all of which (no single determinant here) may structure particular reactions or views about political and social issues. Consider the wealthy who vote for higher taxes on themselves and the working poor who vote for political choices that reduce or eliminate government benefits. Not everyone conforms to group expectations or assumptions. How then can we harness what is good about human diversity to greater mutual empathy and understanding of others who do not share our own views? And, even more importantly, how can we agree to disagree, but still prevent violence, increase human well-being, and adopt some concrete policies?

Consider two recent examples, from two plays that emerged in Los Angeles in the post-Trump era. In *Building the Wall,*[39] it is post-Trump America (2019) and a highly educated African-American history professor, Gloria, is interviewing Rick, a security guard in a privatized prison who is awaiting sentencing for what has come

---

37. *See* AMARTYA SEN, IDENTITY AND VIOLENCE: THE ILLUSION OF DESTINY (2007).

38. *See* LEE ROSS & RICHARD E. NISBETT, THE PERSON AND THE SITUATION: PERSPECTIVES OF SOCIAL PSYCHOLOGY (1991); *see also* THE STANFORD PRISON EXPERIMENT (Abandon Pictures 2015) (depicting Philip Zimbardo's Stanford Prison Experiment); EXPERIMENTER (BB Film Productions 2015) (depicting Stanley Milgram's authority and obedience studies)—both of these studies have been subject to much substantive, methodological and ethical debates for many decades now.

39. ROBERT SCHENKKAN, BUILDING THE WALL (2017) Premiered in Los Angeles, Fountain Theatre with better reviews in LA (Hollywood Reporter) than when it opened in New York, NY Times, May 22, 2017. Playwright Schenkken is also the author of *The Kentucky Cycle* and co-writer of *Hacksaw Ridge.*

to be a mass murder and concentration camp-like carceral state for immigrants, following a national round-up of immigrants which overwhelms the detention centers. While the play builds on eerie resonances to the Holocaust and mimics many actual such interviews by journalists of war criminals,[40] as well as freedom fighters, to attempt to demonstrate, in dialogue, the underlying views of those with such commitments to "causes" (both good and bad), this play demonstrates the tropes of "moral tribes." "No one was speaking for us" says Guard Rick, speaking of his white, working class, Texan, loss of employment "disenfranchised" figure, as if that explains the sliding decay into human depravity and structured loss of life to "those Others." (Speaking to the now well-educated privileged African-American woman who has "triumphed" over an "earlier" racism and injustice in the United States— slavery and discrimination). Though dramatic, the dialogue here does not illuminate much human understanding on either side (where I hoped the play would go) and instead dramatizes the events—this "slippery slope" of Trumpian anti-immigrant feeling will lead the United States into the worse-than-"swamp" of Nazi-Germany with targeted hate and state power.[41] The protagonists argue with each other, describing why they feel the way they do based on their life experiences, but there is no "movement" of mutual recognition. We the audience are to be "shocked" by the revelation that if we don't stop Trump's policies, we will descend into genocides.[42]

In contrast, The *Cake,*[43] based loosely on the legal case now heading to the United States Supreme Court[44] in which a local baker refused to bake a wedding cake for a gay couple on the basis of his "Free Exercise of Religion" (Bible disapproval of gay marriage) under protection of the First Amendment, maintains the political and personal pain of difference across principled divides. A white woman, Jen, who returns home to North Carolina (from NY, where her fiancé, Macy, is a lesbian and African-American,) to ask the white woman (Della) who raised her after her mother's death and who owns a bakeshop, to bake her wedding cake. Despite acknowledging that Jen is like a daughter to her, Della refuses to bake a cake based on her Biblical objections to gay marriage. In a much more subtle dramatization, the parties try to explain their different views, more complicatedly expressed by the

---

40. GITTA SERENY, ALBERT SPEER: HIS BATTLE WITH THE TRUTH (1995).

41. Recently former Governor Arnold Schwarzenegger posted an anti-hate, anti-Trump video decrying the President's failure to detach himself from hate and Neo-Nazis. https://m.youtube.com/watch?v=FN_YIBrOELM (Schwarzenegger, a successful immigrant from Austria, "knew" real Nazis.)

42. As I write this, we seem to have moved someone closer to that mean-spirited dystopia. President Trump has just announced the ending of Deferred Action for Childhood Arrivals (DACA immigration amnesty program created by President Obama as an Executive Order). *See* Michael D. Shear & Julie Hirschfeld Davis, *Trump Moves to End DACA and Calls on Congress to Act,* N.Y. TIMES (Sept. 5, 2017), https://www.nytimes.com/2017/09/05/us/politics/trump-daca-dreamers-immigration.html.

43. BEKAH BRUNSTETTER, THE CAKE Produced and premiered by the Echo Theatre Company, Los Angeles and Ojai, California (2017).

44. *Masterpiece Cakeshop, Ltd. v. Colorado Civil Rights Comm'n.,* 85 USLW 3593 (2012), in which the Colorado Civil Rights Commission held that it was a violation of state anti-discrimination law to refuse to bake a wedding cake for a homosexual couple, whose marriage is now recognized by American Constitutional law, *Obergefell v. Hodges,* 135 S.Ct. 2584 (2015), despite counterclaims by the bakeshop's owner, Jack Phillips, that it is his right, under the First Amendment guarantee of Freedom of Exercise of Religion, to refuse to bake a cake for a wedding he regards as contrary to the Bible and his religious beliefs. *See* Adam Liptak, *Justices to Hear case on Religious Objections to Same-Sex Marriage,* N.Y. Times (June 26, 2017), https://www.nytimes.com/2017/06/26/us/politics/supreme-court-wedding-cake-gay-couple-masterpiece-cakeshop.html.

more radical (also a journalist) New Yorker on the one hand, and the "gentler" version of asking for personal acceptance (not political) from the young woman who returns home to her "adoptive" mother. As the parties engage with each other (and their own sexual and relationship demons), maintaining their clear, strong, emotional and political, if stereotypic, views for most of the play, with Della refusing to bake a cake or go to the wedding, by the end, Macy (who hates sugar), and Della (the baker) find a personal common ground in their love for Jen, which allows Macy to enjoy a sugar cake made, finally, in the bakeshop (not wedding), as both begin to speak across their value differences. This is a more classic play of "resolved conflict" and "changed circumstances" that is the stuff of formal drama (which is as formulaic and structured as our mediation and facilitation "scripts") which demonstrates that time, conflictual conversation,[45] high drama, and yes, direct confrontation (and finally love, not reason) can often lead to "break-through" revelations of human understanding and "acceptance" if not agreement. Drama and emotion, not rational or political argument, bring these people of different value systems together. Whether it will have a lasting impact in either of their differing communities remains to be seen. And, whether anyone who is not already committed to gay marriage will see this play is another issue. Theatre is not the place where complex economic and political policy can be made, of course, even if personal connections can be made.[46] Though, with such issues I am more hopeful—at some point almost every human family will see there is likely a gay person in their midst they love. I can hope that such familial love might warm to or stretch to changed relationships and then, maybe, changed political views.[47] This will not work for all issues of difference, though some day, if not yet, it should work for immigration and interracial and religious marriage too.

---

45. DOUGLAS STONE, BRUCE PATTON, & SHEILA HEEN, DIFFICULT CONVERSATIONS: HOW TO DISCUSS WHAT MATTERS MOST (1st ed. 1999).

46. For arguments that pursuit of the humanities in reading literature and being exposed to art can, in fact, induce more "altruistic" empathy of others, see, for example, MARTHA C. NUSSBAUM, POETIC JUSTICE: THE LITERARY IMAGINATION AND PUBLIC LIFE (1997); *see also* ROBIN WEST, CARING FOR JUSTICE (1999); *but cf.* RICHARD A. POSNER, LAW AND LITERATURE: A MISUNDERSTOOD RELATION (1988) (literature has both good and bad characters and one cannot assume it will make us understand others better or develop empathic understandings that will cause us to behave better).

47. *See* Jennifer Gerarda Brown, *Peacemaking in the Culture War Between Gay Rights and Religious Liberty*, 95 IOWA L. REV. 747 (2010).

IV. HOW WE MIGHT TRY TO GET ALONG: LINKING THE PERSONAL TO THE POLITICAL[48]

A. Getting To Know You

One place to begin is with the processes and protocols now used by many community and policy mediators.[49] No matter how large disputing groups may be (though small is almost always better here) beginning with *personal statements* of who a person is, what the *sources of their identity and beliefs* are, what *major experiences* have molded them (in their own views) and *what concerns or "curiosities" or questions* they have about their own views, often opens up the often hidden assumptions or rigid backgrounds of particular views for further exploration. In such a first stage of engagement the idea is not to challenge the "stickiness" of moral, political, religious, or other values and commitments, but just to listen to their sources as each individual frames them. As Arlie Hochschild describes it "I had a keen interest in how life *feels* to people on the right."[50] Those of us who are such mediators and facilitators use many tools of such human engagement—asking groups of 2's or 3's to share something surprising about themselves (in small groups, rather than large) to broaden a sense of human "knowing" about each other, and often to uncover non-ideological alliances—commonalities around parenting, disability, loss, hobbies, love of animals, shared talents, the arts, sports, and other human connections. The greatest of these forces (and sometimes, of course, also the most destructive), is cross-group love; think Romeo and Juliet, Tony and Maria in *West Side Story,* and more recently, the doomed lovers in Bosnia in Angelina Jolie's *The Land of Blood and Honey.*[51]

These activities are not only instrumental "ice breakers," they are intended to form *bridges* of human connections and perhaps some trust across the very values that divide us. This is important for the one instrumental process that later can be used to bridge differences: *bargaining and trades,* and some compromise, in a positive way.[52] Excavating sources of beliefs and commitments can be both troublesome (crystallizing and hardening such commitments) and an opportunity for exploration, exposing the possibilities that even if there is never agreement on big picture "worldviews," there can sometimes be agreement on more concrete and practical issues (e.g. abortion dialogue leading to practices regarding clinic protests

---

48. This old political slogan belongs to the feminist movement demanding political recognition for legal and other social issues located in the private sphere and family life, but here I recall it to make the point that political disagreements of all kinds are now experienced on a personal, affective level and appeals to reason alone, assuming cognitive or conceptual attention to issues that people have strong "feelings" about will not suffice to encourage engagement across political and policy divides. I am not saying that "facts don't matter," but rather that they are not enough alone. (Note how effective President Ronald Regan was in attracting support by using "personal anecdotes".)

49. *See* PODZIBA, *supra* note 15; ESSENTIAL PARTNERS, About Us, https://www.whatisessential.org/about-us (last visited Nov. 13, 2017).

50. HOCHSCHILD, *supra* note 10, at ix.

51. Carrie Menkel-Meadow, *In the Land of Blood and Honey: What's Fair or Just in Love and War Crimes? Lessons for Transitional Justice, in* FRAMING LAW AND CRIME: AN INTERDISCIPLINARY ANTHOLOGY (Caroline Joan S. Picart, Michael Hviid Jacobsen & Cecil Greek, eds., 2016).

52. Carrie Menkel-Meadow, *The Ethics of Compromise, in* GLOBAL ENCYCLOPEDIA OF PUBLIC ADMINISTRATION, PUBLIC POLICY AND GOVERNANCE (Ali Farazmand, ed., 2016).

and safety,[53] HIV-AIDS policies,[54] environmental siting and policy,[55] and perhaps in present times disaster (hurricane) relief, re-building infrastructure in current US, and maybe even Congressional action on DACA reform and renewal).

The protocols developed by such groups as Public Conversations, now Essential Partners, and other community and public policy facilitators,[56] offer guidelines for having parties in dispute and conflict approach each other with human respect and mutual curiosity to consider such questions, with respect to any disputed issue:

1. From where do your views on that topic come? (your personal experience, education, religion, profession, community, etc.*);*

2. What forces/people influence your views? (opinion leaders, clergy, family, colleagues, professional community);

3. What questions (points of curiosity) do you have for others who have different views?

4. What doubts or uncertainties do you have about your own views?

5. What other information would you need/like to have to answer some of these questions?

B. Engaging on all Cylinders: Empathy

Responses to and engagement with these questions include facts, arguments, belief systems, personal experiences and acknowledgement of the relational nature of knowledge and understanding. When discussed, in facilitated and structured dialogic formats, the answers to these questions involve aspects of cognitive knowing (brain), affective knowing (heart) and even physical knowing (stomach) as people relate their personal histories and their relation to broader worldviews and opinions on particular issues. Slow and skillful management of guided conversations around these questions in different size groups on different issues often exposes the sources of people's views that come from different life experiences and sources, and can permit human questioning and curiosity to see others from where they are literally "coming from." Most of us have experienced those moments of emotional recognition—empathy—that allows us to "be" with someone's else's reality, even if we don't agree with them and don't share their life experiences.

---

53. *See* PODZIBA, *supra* note 15, at 75-96; *see also* Michelle LeBaron & Nike Carstarphen, *Finding Common Ground on Abortion, in* CONSENSUS BUILDING HANDBOOK 1031 (Carrie Menkel-Meadow, commentator at 1034-1050).

54. *See* John Forester, *Dealing with Deep Value Differences, in* CONSENSUS BUILDING HANDBOOK 463; *see also* Michael Hughes et al., *Facilitating Statewide HIV/AIDS Priorities in Colorado, in* CONSENSUS BUILDING HANDBOOK 1011.

55. *See* Judith Innes & Sarah Connick, *San Francisco Estuary Project, in* CONSENSUS BUILDING HANDBOOK 801; *see also*Edward Scher, *Negotiating Superfund Cleanup at the Massachusetts Military Reservation, in* CONSENSUS BUILDING HANDBOOK 859.

56. *See* International Association for Public Participation, http://www.iap2.org (last visited Nov. 13, 2017) (which provides protocols, assistance and leadership on facilitating community, group and political meetings to manage such proceedings)

As I tell my clients and students, to "walk a mile in another's shoes" may give you some feelings of *sympathy* (you are still using your own feet [read own values to filter your experience]), but you need to "walk a mile with the other person's feet" (their values) in order to experience true empathy. "I feel your pain" can be heard as patronizing to one who has not lost a loved one, felt actual discrimination, been in a car crash or been deported, so more authentic experiences and words of empathic understanding must be found for listening to other's sources of information and belief. Consider President Obama's famous race speech in which he acknowledged his own experience of both his black and white parents and grandparents, and his own "cringing" at racist statements made by his white grandmother, but also his discomfort with some of the racist statements made by some of his black mentors and clergy.[57] Honest acknowledgement and self-confrontation about complex and very sensitive issues brings difficult issues to the fore and suggests what we can learn from how complicated and *different* our own experiences can be. President Obama, as a modern mixed race individual, has the human advantage of actually being able to experience "both sides now"[58] from inside and out. In a peace project I have worked with in Israel-Palestine, attempts to get parties on different sides of the conflict to be empathic with each other failed when Palestinians did not see what Yad Vashem (Museum of the European genocide of the Jews) had to do with their own pain in their diaspora from their displacement after the Israeli war of Independence and removal from their own homeland. Empathy is not false equivalence—it is a form of "feeling into" the "other" (from the German – *einfühlung),* which is one of the sources of the modern use of the word, derived from the Greek pathos (feeling, emotion). It is a state of understanding others "from their own frame of reference."

The use of these questions to broaden out conflictual positions is not to reach "equivalence" of views or necessary agreement on ultimate truth, but rather to really "understand" where the other person or "side" is coming from in order to develop questions, agendas, and searches for new insights, information and only later, and not in all cases, some concrete actions. When used in complex disputes such personal engagement can be followed with fact-gathering, agreements to pursue joint expert or science panels[59] for more information, followed by intensive and structured bargaining to reach concrete, but often *contingent* agreements on particular issues or tasks.

---

57. Senator Barack Obama, A More Perfect Union (May 18, 2008), (transcript available at http://www.nytimes.com/2008/03/18/us/politics/18text-obama.html).

58. JONI MITCHELL, BOTH SIDES NOW (Reprise 1969) (my favorite version was sung by Judy Collins on her album WILDFLOWERS, released in 1967) (my anthem for life and mediation!) "I've looked at life, love and clouds from both sides now, from win and lose, but I still really don't know life/love/clouds at all" (curiosity and continued engagement anyone?).

59. See, for example, Gina Kolata, *Panel Can't Link Breast Implants To Any Diseases,* N.Y. TIMES (Dec. 2, 1998), http://www.nytimes.com/1998/12/02/us/panel-can-t-link-breast-implants-to-any-diseases.html, for the use of the Expert Epidemiology Panel in the Silicon Breast Implant Litigation appointed by federal Judge Sam Pointer to resolve the science in the mass tort class action involving breast implants and auto-immune disease.

Different structures and processes which use these techniques have been used in empathy training in divorce mediation and collaborative law,[60] personal relations,[61] defusing and managing community disputes,[62] as well as political disputes within and between nations[63] (often used intuitively by the best of international peace mediators).[64] The challenge is to permit enough time for the sharing and "processing" of roots and grounds (sources) of people's ideas and commitments, with appropriate acknowledgement and respect,[65] and then to "move forward" (not "move on," which is dismissive), with this knowledge and emotional engagement, to more concrete attempts, where possible, to work out, bargain about and trade (negotiate and sometimes compromise productively) particular elements to create some concrete agreement or action. As one of our most gifted practitioners and theorists, John Paul Lederach, has suggested, to "transform" a conflict one has to want to create a better state of the world, from sitting in the middle of the conflict and "bad" state. In retrospect, many have credited sage leaders as those who were able to "imagine" (with empathy) how their counterparts were thinking, from their (the counterpart's) value systems, to strategize, negotiate, and act. See reports of John F. Kennedy and Robert Kennedy in the Cuban Missile Crisis[66] and Jimmy Carter at Camp David.[67] These leaders knew what would "work," with their negotiating partners—threats, but also more palatable intermediate steps, concessions, and agreements (e.g. removal of US missiles from Turkey and "neutralized return" of the Sinai to Egypt) —because they imagined how the "other" leader would have to respond to their own constituent's needs and interests (and emotions!). How individuals, groups, organizations, parties, and nations can move themselves from an undesired state of conflict and pain to a more desirable (and achievable) state (even if temporary, or contingent) often requires a skillful manager of process, and transition from an individualized process of emotional sharing to a more functional and practical problem-solving (trading, bargaining) and instrumental process of guided negotiation.

---

60. *See generally* GARY FRIEDMAN & JACK HIMMELSTEIN, CHALLENGING CONFLICT: MEDIATION THROUGH UNDERSTANDING (2008).

61. *See generally* MARSHALL B. ROSENBERG, NON-VIOLENT COMMUNICATION (3d ed. 2015); *see also* SAM HORN, TONGUE FU: HOW TO DEFLECT, DISARM, AND DEFUSE ANY VERBAL CONFLICT (1996).

62. *See generally* JAY ROTHMAN, RESOLVING IDENTITY BASED CONFLICTS IN NATIONS, ORGANIZATIONS AND COMMUNITIES (1997); see also Nancy H Rogers, *When Conflicts Polarize Communities: Designing Localized Offices That Intervene Collaboratively*, OHIO ST. J. OF DISP. RESOL. 2015; Lela P. Love, *Glen Cove: Mediation Achieves What Litigation Cannot," in* STORIES MEDIATORS TELL 117 (Eric R. Galton & Lela Love, eds., 2012).

63. JOHN PAUL LEDERACH, BUILDING PEACE: SUSTAINABLE RECONCILIATION IN DIVIDED SOCIETIES (1997); JOHN PAUL LEDERACH, THE LITTLE BOOK OF CONFLICT TRANSFORMATION (2003).

64. *See, e.g.*, GEORGE J. MITCHELL, MAKING PEACE *(1999)*.

65. *See* Carrie Menkel-Meadow, *Remembrance of Things Past? The Relationship of Past to Future in Pursuing Justice in Mediation*," 5 CARDOZO J. OF CONFLICT RESOL. 97 (2004).

66. *See* ROBERT KENNEDY, THIRTEEN DAYS: A MEMOIR OF THE CUBAN MISSILE CRISIS (1969); see also DEEPAK MALHOTRA, NEGOTIATING THE IMPOSSIBLE: HOW TO BREAK DEADLOCKS AND RESOLVE UGLY CONFLICTS (WITHOUT MONEY OR MUSCLE) (2016) (see Ch. 13, "The Power of Empathy" reporting JFK's empathic responses to Nikita Khrushev's political dilemmas).

67. *See* LAWRENCE WRIGHT, THIRTEEN DAYS IN SEPTEMBER: THE DRAMATIC STORY OF THE STRUGGLE FOR PEACE (2014) (recounting the peace negotiations at Camp David 1979-80); *see also* JIMMY CARTER, KEEPING FAITH: MEMOIRS OF A PRESIDENT (1995).

C. Contact Theory and Perspective Taking

Research in the behavioral sciences and some evaluation of work done by public policy mediators suggest some fruitful avenues to pursue and also some pitfalls of which to be aware. In sociology and social psychology, "contact theory" has long posited that increased "positive" contact opportunities between distrusting or conflictual groups can improve inter-group attitudes and behavior.[68] Researchers in post-conflict, reconciliation-seeking societies (Colombia, South Africa, Israel-Palestine, Cambodia, Argentina, Chile,[69] and others) are now producing mixed evaluations of how such structured contacts—truth and reconciliation events, and various forms of narrative "perspective taking" events—are working.[70] Such processes entail psychological and social processes (heart, stomach, and brain) as painful experiences are recounted and "repeated" for understanding (active listening and other techniques) which are intended to "reframe" cognitive understandings of past events, "heal" emotional harm, and "reorient" parties to each other, sometimes with the hope of creating "new" narratives to accompany, if not supplant, older narratives of conflict and distrust. At a political level, such "encounters" have more instrumental goals—to form new group, national and even individual identities, and to create new policies (e.g. the South African constitutional process),[71] legislation and action plans for projects at various levels.

In political, post-conflict restorative work, exercises in "perspective taking," narrative and dialogue groups, and reframing in social and political terms (formal role-reversal in mediation or truth and reconciliation settings) is sometimes effective as an emotional exercise before the hard work of bargaining or co-existence can occur, but this too can be dangerous and make things worse, if not well-managed. In some highly conflictual settings, power sharing (from compromise processes) may be more effective (as in Northern Ireland) or formal division (think Cyprus or Virginia and West Virginia before the Civil War). The research arm of the conflict resolution and peace studies field is now busily studying, in many venues, how different contact protocols and peace interventions are working around the world.[72] We should have many opportunities for case studies here in the United States.

---

68. *See* GORDON W. ALLPORT, THE NATURE OF PREJUDICE (1979); *see also* T.F. Pettigrew & L.R. Tropp, *A meta-analytic test of intergroup contact theory*, 90(*J. OF PERSONALITY AND SOC. PSYCHOL.* 751 (2006).

69. Carrie Menkel-Meadow, *Process Pluralism in Transitional/Restorative Justice: Lessons from Dispute Resolution for Cultural Variations in Goals beyond Rule of Law and Democracy Development (Argentina and Chile)*,3 INT'L J. OF CONFLICT ENGAGEMENT AND RESOL. 3(2015).

70. *See, e.g.,* Juan E. Ugarriza & Enzo Nussio, *The Effect of Perspective-Giving on Postconflict Reconciliation, An Experimental Approach*, 38 POL. PSYCHOL. 3 (2017); Daniel Bar-Tal & Gemma H. Bennink, *The Nature of Reconciliation as an Outcome and as a Process, in* FROM CONFLICT RESOLUTION TO RECONCILIATION 11 (Yaacov Bar-Siman-Tov ed., 2004).

71. *See, e.g.,* James Gibson, *Does Truth Lead to Reconciliation? Testing the Causal Assumptions of the South African Truth and Reconciliation Process*, 48 AM. J. OF POL. SCI. 201 (2004); *see also* THE POST-APARTHEID CONSTITUTIONS: PERSPECTIVES ON SOUTH AFRICA'S BASIC LAW (Penelope Andrews & Stephen Ellmann eds., 2001).

72. *See, e.g.,* CRAIG ZELIZER & ROBERT A. RUBINSTEIN, BUILDING PEACE: PRACTICAL REFLECTIONS FROM THE FIELD (2009); *see also* . HERDING CATS: MULTIPARTY MEDIATION IN A COMPLEX WORLD (Chester A. Crocker, Fen Osler Hampson & Pamela Aall eds, *1999*).

D. Moral and Political Reframing

Recent work in the field of moral psychology suggests that a sort of "moral reframing" of arguments (a mixture of cognitive and affective appeals) on particular issues to appeal to the values of "the other" (e.g. using conservative versions of arguments for health care, immigration, tax reform), rather than to more arguments of one's own (more facts and information from a "liberal" perspective) can be more effective for acceptance of particular policy ideas than simply recasting arguments with more information or data from one's own side.[73] This form of "reframing" is both "effective," but also potentially objectionable to those who feel it is a "compromise" of one's own values. For me, this is both a question of desired instrumentalism and expediency, as well as testing our moral limits.[74]

E. Bargaining and Compromise

Some "compromise" is absolutely essential (and moral, in my view) to get things done. In the words of the United States Magistrate Judge (who first inspired my work) in a federal lawsuit settlement conference, "to sue (fight) is human, to settle divine."[75] For those who work with the Thomas-Kilman MODE instrument, the choice for one conflict strategy (in some settings) to "I will let him/her have some of his/her views, if s/he will let me have some of mine"[76] expresses how political agreements can be made to accomplish legislation and other arrangements, without full agreement on underlying values. Most tax legislation reflects such "compromise," as has immigration and health care regulation. In ordinary contract formation, as well as treaty negotiation, studied ambiguity and a failure to fully agree on the actual meaning of all terms and goals and purposes of such documents, reflects the fact that sometimes having an agreement is better than none at all. This often puts off to a later time particular conflicts about interpretations, by avoiding detailed specification of ultimate purposive and value-based meaning. This is the middle column of my chart above—between principles (reason) and values (heart) we use bargaining and trades (and our "guts") to get things done—stop the violence, provide disaster aid, raise money for the common good, and pay for our public services, all while still arguing about "optimal" health, education, and foreign policy.

As any negotiation teacher and practitioner well knows, the more issues the merrier. The fact that we have different values or different desiderata is a good thing for trading and bargaining. Yes, mediators ask us to look for "common ground," but actually I am most often interested in non-competing "complementary" goals (I like the icing or orange peel, you like the cake or juice). Where, even with different reasons or values, can we get something done (shared, expandable, transformable resources, items and yes, even views). Della doesn't have to come to a wedding she disapproves of, but if she makes a cake for her lesbian "daughter" and partner she may come to appreciate their relationship. And you can be sure if

---

73. Matthew Feinberg & Robb Willer, *From Gulf to Bridge: When Do Moral Arguments Facilitate Political Influence?*, PERSONALITY & SOC. PSYCHOLOGY BULL. 1665 (2015)

74. See my work on "compromise", Menkel-Meadow, *supra* note 49.

75. *See* Menkel-Meadow, *supra* note 26, at 754.

76. *See* Kenneth Thomas, *Conflict and Conflict Management, in* HANDBOOK OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY (Marvin D. Dunnette ed., 1976).

the couple "expands" their family to children, Della is very likely to want a relationship with her grandchild. I have seen it happen—with gay, inter-racial, inter-class, and inter-religious couples and families. My German refugee father, who served in the United States Army in World War II in Hawaii, strongly believed the answer to all racism and intolerance was intermarriage. The US Census now has to deal with just how much of this is happening.[77] Observing how all the "modern families" are negotiating their differences (in real life, as well as on TV and in the movies—see *Loving*[78] and *This Is Us*[79]), we can see how daily life is negotiated across all spheres, directly and often with a lot of legal and emotional conflict.

F. *Expert Facilitation: A Democratic/Legitimacy Issue?*

The design and management of such processes raise an important and ironic question for conflict resolution and deliberative democracy—the *need for expertise and facilitation,* in what many democratic theorists suggest may be a distortion in purely democratic processes (e.g., the legitimacy of unelected leaders, leaders at all). Our field has considered these issues before and recent critiques of leaderless political movements (think Occupy Wall Street[80]) suggest that many have come to appreciate the value of our profession—groups and individuals in conflict need help to sort out their differences politically, socially, and emotionally. Our own Constitutional formation processes benefitted greatly from the process expertise of George Washington (leader), James Madison (recorder and agenda manager), and Benjamin Franklin (affective political facilitator and observer). [81] And the contact theory, discussed above, which reports on conditions under which positive reinforcement and reconciliation may occur, documents that such encounters are more successful when conducted in the presence of an external and skilled outsider.[82] So, we should explore the technical aspects of our craft and also delve deeply into the legitimacy of "consent" and other justifications for such work.

Using the tools of multiple "consciousnesses" reviewed here (the rational-principled, mediated and negotiated bargaining, and appreciation of affective and moral commitments), professional facilitators and mediators can probe substantive facts (and science and data!) like experts or investigatory bodies,[83] counsel and guide bargaining and trades based on needs and interests, and provide structured engagement over difficult emotional, moral, religious and value based commitments (as Howard Bellman says, "having an ear").[84] Modern forms of facilitated dialogue (unlike debates or legislative votes or executive decrees) do allow all of these levels to be explored in the same room. If there is negotiated agreement on procedural

---

77. See generally Naomi Mezey, *Erasure and Recognition: The Census, Race and the National Imagination,* 97 NW. U. L. REV. 1701 (2003).

78. LOVING (Raindog Films 2016).

79. *This is Us* (NBC Television Broadcast 2016).

80. Nathan Heller, *Out of Action: Do Protests Work?,* THE NEW YORKER 70, 77 *(*Aug. 21, 2017).

81. For a masterful description of our Constitutional framing process as an exercise of facilitated negotiation, see Dana Lansky, *Proceeding to a Constitution: A Multi-Party Negotiation Analysis of the Constitutional Convention of 1787,* 5 HARVARD NEGOT. J. 279 (2000); for a fuller historical elaboration of the process rules, deliberations and complex interest group negotiations, see MICHAEL J. KLARMAN, FRAMER'S COUP: THE MAKING OF THE UNITED STATES CONSTITUTION (2016).

82. Ugarriza & Nussio, *supra* note 70, at 5.

83. *See* PODZIBA, *supra* note 15, at 178-79.

84. *Id.* at 180.

ground rules and decision rules, such groups have been more likely to transcend differences to accomplish effective decision making, in both big matters (abortion, AIDS policy, animal rights)[85] and more local (land-use, zoning, community relations).[86] Voting rules, of course, matter enormously and we should all be experts in the effects of different voting choices.[87] Many (non-legislative) situations don't even require formal votes, as some settings may be better serialized for different meetings for learning and engagement, before deciding anything. Acknowledging that there are great variations within groups (Conservative Libertarians [think William Buckley] sometimes have more in common with some Liberals—free speech and press, government out of social life) allows incremental or smaller agreements to occur or "cross alliances" on particular issues. Consider Republican Conservative John McCain's role in anti-torture legislation, supported by Democrats, because of his own personal experience as a war prisoner. Legislation, deals, treaties, contracts and relationships are all made when we each get something of what we want, in return for the others getting some of what they want. Sometimes agreement on *why* we want those things may not even have to be resolved. Consider, for the moment, the few big issues we could possibly make some progress on now, even given the totally polarized nation we inhabit—a public-private partnership on infra-structural building and repair; immigration policy (at this writing, some renewal of DACA); potential amendments to the Affordable Care Act; tax reform; emergency disaster relief...[88]

Being heard at all these levels, being gently, but acutely, questioned about where one's facts and assumptions come from, often allows new ideas to emerge in "safe zones" for trying out of new contingencies and solutions, and, as mediators say, "reality testing" of ideas that may be counter to the old ways of doing things or simply different from the assumptions or starting points of a particular dispute.

G. Can We Get Beyond Reactive Devaluation?

More difficult, of course, is dealing with groups (like current "reds" or "blues") who assume so much value difference or conflict (e.g., identity conflict) that they cannot credit anything offered by "the other side," known to us as "reactive devaluation" bias.[89] When we automatically discount or don't even hear others, any engagement may be impossible. Enter the mediator to reframe proposals, ideas and arguments in more "neutral" terms so the parties can then evaluate without attribution to a particular "enemy". Other tools available for trust building—or as one of

---

85. JEFFREY CRUIKSHANK & LAWRENCE SUSSKIND, BREAKING THE IMPASSE: CONSENSUAL APPROACHES TO RESOLVING PUBLIC DISPUTES (1987).

86. *See, e.g.,* Susskind, et. Al., CONSENSUS BUILDING HANDBOOK, *supra* note 15, Cases 1, 4, 10, 11, 12.

87. Carrie Menkel-Meadow et al., DISPUTE RESOLUTION: BEYOND THE ADVERSARIAL MODEL 581-627 (2nd ed. 2010); LAWRENCE SUSSKIND & JEFFREY CRUIKSHANK, BREAKING ROBERT'S RULES: THE NEW WAY TO RUN YOUR MEETING, BUILD CONSENSUS AND GET RESULTS (2006).

88. At this writing President Trump has agreed with Democrats to raise the debt ceiling, in light of the vast amount of federal aid that will be needed to deal with two enormous Hurricanes (Harvey and Irma) affecting two (Republican) states—Texas and Florida. See Peter Baker et al., *Trump Bypasses Republicans to Strike Deal on Debt Limit and Harvey Aid,* N.Y. TIMES (Sept. 6, 2017), https://www.nytimes.com/2017/09/06/us/politics/house-vote-harvey-aid-debt-ceiling.html.

89. Lee Ross, *Reactive Devaluation in Negotiation and Conflict Resolution, in* BARRIERS TO CONFLICT RESOLUTION 26 (Kenneth J. Arrow et al. eds., (1995).

my mentors, gestalt therapist Janet Lederman, used to say, "get the focus off the focus"—involve disputing parties in a joint activity or task that has nothing to do with their conflicts (think of the cross-group alliances after disasters—in hospitals, playing with children, charitable arts, and sports events, etc.). All of the intellectual and social energy that has gone into restoring some of the "lost" social capital observed by Robert Putnam and his colleagues in *Bowling Alone*,[90] by encouraging cross-class, interfaith, citizen dialogue, and other social groupings is an effort to "return" to a society of great complexity and diversity with more locations of intergroup interaction. Recent Pew surveys and other data[91] seem to suggest that we are in fact living further and further apart from those who are different from us, so that it is perhaps only the workplace, and for some of us still, public education, where we can truly encounter each other to try out these many suggestions for better engagement. But I continue to worry.

V. A FEW CONCERNS: WHEN AND WHY WE SOMETIMES JUST CAN'T GET ALONG

I have spent a good part of my professional life as an educator and mediator, in both roles acting from a place where I believe that people can learn new things, analyze old things, and come to new understandings of how to make the world a better place. I have facilitated hundreds, probably thousands, of classes, lawsuits, disputes, and meetings in which people of different values, information and views have come together in some common enterprise (by definition they have entered the same room, even if the room is only virtual these days) to listen to each other and often, to do something—settle a lawsuit, improve a relationship, reconfigure an organization or make a strategic plan. Never since I have been on this earth have I been so discouraged about our current polity. (I was born in the same year and right next to the United Nations, and grew up with the children of its diplomats; my parents were refugees from Hitler's Germany, so we were optimistic cosmopolitans, hoping that the world could be made a better place.) Unfortunately, Donald Trump grew up very near me in New York and was already a bully when he was a very young man.[92] For the first time in recent history, we have a president with no experience in governmental discourse, deliberation, purpose, and management. Even Republican presidents who wanted less government did incredible things (think Nixon founding EPA and opening relations with China). I watched the "take no prisoners" Gingrich Revolution destroy cordial relations in the US Congress and Washington D.C. generally, even as skilled facilitators tried to create a "harmony" retreat (in Hershey, Pennsylvania) for members of Congress, but things have never been the "same" since.

All that I (and many others) have described here—using empathy, personal stories and narratives, data gathering, curiosity sharing, reality testing—all of these things require good will and trust and a willingness to be in the same room! As

---

90. ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY (1st ed. 2001).

91. *See* Charles Babington, *A Polarized America Lives as It Votes*," TRUST MAG., (Sept. 1, 2014*),* http://magazine.pwetruts.org/en/archive/summer-2014/a-polarized-america-lives-as-it-votes.

92. *See* Donald J. Trump & Tony Schwartz, TRUMP: THE ART OF THE DEAL (1987); *see also* Jane Mayer, *Donald Trump's Ghostwriter Tells All*, THE NEW YORKER (July 25, 2016), http://www.newyorker.com/magazine/2016/07/25/donald-trumps-ghostwriter-tells-all.

others have explored when we "should not bargain with the devil,"[93] I now wonder whether we can truly talk and hear each other across such great value divides. When a President says there were some "very good people" among Neo-Nazis and race-baiters and haters, I wonder where our national good will has gone. I have been an "unbiased" mediator and facilitator, as well as a protest marcher and political activist, an educator and a governmental consultant, always looking for common "human" ground, across differences, and enjoying both human commonalities and shared destinies, and also reveling in our human differences while learning from those who are different from me. I do experiments in my head to try to imagine how Donald Trump and some of his supporters would do in a "healing circle," negotiated rule-making, facilitated strategic planning exercise for the West Wing, family mediation, or simply a "brainstorming" Cabinet meeting, and I can't even conjure up a good fantasy movie. As Charles Dickens famously said, "it was the worst of times" (I am leaving out "the best of times"[94]), for us, at least at the national policy implementation level.

But, I remain somewhat optimistic, that in the interstices of federal, and many state, agencies and offices, and in private work settings, universities and organizations, people who care about each other will use the techniques of conflict resolution and sensible policy management to continue to set the table, sow the land, clean the machinery, and practice their scales to keep ourselves ready, not rusty, to work wherever we can—to keep doing facilitation, empathy trainings, personal narrative workshops, consensus building exercises, mediated negotiation, (and for me, teaching and working with my students on our annual Global Justice Summit) in order to innovate new policy solutions to seemingly intractable problems, provided the weather and geo-politics allow it.

Inside of cursing the darkness, I will light a candle and ask a question of curiosity. I hope you will too.

---

93. ROBERT MNOOKIN, BARGAINING WITH THE DEVIL: WHEN TO NEGOTIATE, WHEN TO FIGHT *(2010)*; AVISHAI MARGALIT, ON COMPROMISE AND ROTTEN COMPROMISES (2010).
94. CHARLES DICKENS, A TALE OF TWO CITIES (1859).

BUILDING CONFLICT RESILIENCE: IT'S NOT *JUST* ABOUT PROBLEM-SOLVING

Robert C. Bordone*

I. INTRODUCTION

Political polarization in the United States and internationally has increased enormously in the past decade, resulting in legislative impasse in some countries, political instability and partisan re-alignment in others, and decreased levels of communication, trust, and cooperation across partisan lines in schools, communities, and across the nation.[1] Americans are more likely to marry someone of a different faith than of a different political party, a statistic that would have been unheard of a generation ago.[2] In the United States, political polarization has made effective governance nearly impossible, resulting in an inability of political leaders to communicate constructively with each other even on some of the nation's most pressing problems, such as healthcare and immigration.[3] Virtually every issue in American politics is framed in zero-sum, winner-takes-all terms.[4]

Many factors have brought us to this moment of political polarization after several decades of broad national consensus in the U.S following World War II.[5] Among the factors contributing to polarization are highly partisan cable news networks;[6] social media such as Facebook and Twitter, which creates political echo chambers and allow users to curate whom and what they read and pay attention to;[7] and the rise of anonymous blogs that stifle in-person conversations for fear that

---

* Thaddeus R. Beal Clinical Professor of Law, Harvard Law School. The author extends a special thanks to June Casey, Research Librarian at Harvard Law School, for her research assistance and enthusiasm for this project.

1. *See, e.g.*, Hanno Sauer, *Can't We All Disagree More Constructively? Moral Foundations, Moral Reasoning, and Political Disagreement*, 8 NEUROETHICS 153, 153 (2015).

2. *Inter-Faith Marriage: Across the Aisles—A Welcome Sign of Tolerance, or Dangerous Dilution?*, THE ECONOMIST (Apr. 10, 2013), https://www.economist.com/blogs/democracyinamerica/2013/04/inter-faith-marriage (last visited Sept. 5, 2017).

3. *See* Sarah Binder, *How Political Polarization Creates Stalemate and Undermines Lawmaking*, WASH. POST (Jan. 13, 2014), https://www.washingtonpost.com/news/monkey-cage/wp/2014/01/13/how-political-polarization-creates-stalemate-and-undermines-lawmaking/?utm_term=.260bf8e69393 (stating that legislative deadlock increases as party polarization and partisanship does); *see also* Michael Barber & Nolan McCarty, *Causes and Consequences of Polarization*, *in* NEGOTIATING AGREEMENT IN POLITICS, 19, 39-41 (Jane Mansbridge & Cathie Jo Martin eds., 2013) (using immigration as a case study in legislative gridlock caused by political polarization).

4. Barber & McCarty, *supra* note 3, at 45.

5. Godfrey Hodgson, *Revisiting the Liberal Consensus*, *in* THE LIBERAL CONSENSUS RECONSIDERED: AMERICAN POLITICS AND SOCIETY IN THE POSTWAR ERA 12-16 (Robert Mason & Iwan Morgan eds., 2017).

6. Carl M. Cannon, *Comment*, *in* RED AND BLUE NATION?: CHARACTERISTICS AND CAUSES OF AMERICA'S POLARIZED POLITICS 163, 166-67 (Pietro S. Nivola & David W. Brady, eds. 2006).

7. *Id.*

what they say may be taken out-of-context and posted publicly by others anony-
mously.[8] In the United States, the role of money in political campaigns and the
impact of election gerrymandering have compounded the problem of polarization
whereby ideologically-motivated individuals and strong partisans on both sides
have an outsized impact on policy-making, governance, and the national conversa-
tion while ordinary citizens disengage and often feel disgusted by the decreasing
civility of political elites.[9] The situation has become so dire that the Stavros Niar-
chos Foundation recently made a $150 million gift to Johns Hopkins University in
an effort to identify and study ways to stem the deterioration of civic engagement
and restore civil civic discourse.[10]

II. THE TROUBLING DECLINE OF POLITICAL DIALOGUE IN LAW SCHOOL CLASSROOMS AND BEYOND

As partisan polarization has increased, I have observed several perceptible
changes in the way law school students engage each other around political differ-
ences and conflict in the classroom. From conversations with my colleagues, I am
not alone in observing these trends.
     There has been a noticeable decline in robust exchange around in-class discus-
sion of politically controversial topics such a gender, violence, or rape, for exam-
ple.[11] A dominant political orthodoxy—often left-leaning—rules in many law
school classrooms such that conservative perspectives and more arguably extreme
views on the far left often do not get voiced.[12] This observation would matter less
if there were evidence that there were no students in law school classrooms who
held such views. Unfortunately, this does not seem to be the case.[13] There is ample
evidence to suggest that students with conservative or with far-left views are well-

---

8. Maria Konnikova, *The Psychology of Online Comments*, THE NEW YORKER (Oct. 23, 2013),
https://www.newyorker.com/tech/elements/the-psychology-of-online-comments.
    9. JENNIFER L. LAWLESS & RICHARD L. FOX, RUNNING FROM OFFICE: WHY YOUNG AMERICANS
ARE TURNED OFF TO POLITICS 7-16 (2015); Lawrence Lessig, *In Washington, Money Talks Louder than
Ordinary Americans—and We Do Nothing*, NEWSTATESMAN (June 4, 2015), https://www.newstates-
man.com/2015/05/washington-money-talks-louder-ordinary-americans; Russell Berman, *What's the
Answer to Political Polarization in the U.S.?*, THE ATLANTIC (Mar. 8, 2016), https://www.theatlan-
tic.com/politics/archive/2016/03/whats-the-answer-to-political-polarization/470163/.
    10. *Johns Hopkins Gets $150M for Interdisciplinary Effort to Foster Discussion of Divisive Issues*,
JOHNS HOPKINS UNIVERSITY HUB (June 22, 2017), https://hub.jhu.edu/2017/06/22/agora-institute-150-
million-grant-stavros-niarchos-foundation/.
    11. Jeannie Suk Gerson, *The Trouble with Teaching Rape Law*, THE NEW YORKER (Dec. 15, 2014),
https://www.newyorker.com/news/news-desk/trouble-teaching-rape-law; *See also* Jamie R. Abrams,
Experiential Learning and Assessment in the Era of Donald Trump, 55 DUQ. L. REV. 75, 87 (2017) (not-
ing that law students and faculty can isolate themselves into "political bubbles" and "ideologically ho-
mogeneous environments").
    12. *See, e.g.*, James C. Phillips, *Why are There so few Conservatives and Libertarians in Legal Aca-
demia? An Empirical Exploration of Three Hypotheses*, 39 HARV. J.L. & PUB. POL'Y 153, 159 (2015)
(noting that the large majority of legal academics are liberal in orientation).
    13. Oliver Roeder, *The Most Conservative and Most Liberal Elite Law Schools*, FIVETHIRTYEIGHT
(Dec. 5, 2014, 6:00 AM), https://fivethirtyeight.com/features/the-most-conservative-and-most-liberal-
elite-law-schools/ (demonstrating a high number of both liberal and conservative law clerks and thereby,
implicitly suggesting that student bodies are comprised of both liberal and conservative students); Steven
C. Bahls, *Political Correctness and the American Law School*, 69 WASH. U. L. Q. 1041, 1043-48 (1991)
(citing to responses from an ABA survey of law students, over half of whom do not feel free to express
disagreement with the political perspectives of their professors).

represented on law school campuses.[14] Instead, several phenomena—perhaps most notably the worry that a misunderstood, unpopular, or inartfully worded comment or idea might get posted on a Facebook wall or an anonymous blog in a way that mischaracterizes or labels the speaker—has encouraged many to remain safe by remaining silent. The risk that a poorly phrased statement or unpopular political view could be easily mischaracterized and suddenly posted on a range of public sites thereby casting the speaker in an unfavorable (and perhaps wholly unfair) light to the broader community is simply too high for many.[15]

With political dialogue so charged, students and faculty opt to remain quiet and share their views outside the classroom only with those who are known to be politically like-minded and sympathetic.[16] This trend is troubling in any context, but it is particularly worrisome in an academic environment where expression of a wide range of views and political opinions should be most protected, encouraged, and permissible. As admissions offices work tirelessly to recruit diverse student bodies for the purpose of promoting a wide range of world views and opinions both in and outside the classroom, powerful forces in the society simultaneously raise the cost of genuine face-to-face political exchange with those who have views different from one's own or from the prevailing orthodoxy of a particular community.

These observable trends in the law school classroom reflect the growing polarization in communities, churches, civic organizations, and even families.[17] Holiday get-togethers with family members, fraught with the possibility of political clashes, are frequently managed with an unspoken détente, a tacit agreement to avoid politics all-together as the family eats the turkey and stuffing.[18] These strategies of full-on political clash or nearly complete avoidance should be deeply troubling to those of us who teach, practice, and write about conflict management. Fight or flight are

---

14. Jonas Blank, *All the Right's Moves*, HARVARD LAW TODAY (Apr. 24, 2003), https://today.law.harvard.edu/feature/rights-moves/; *Yale Federalist Society: Lawyering Up Right*, YALE LAW SCHOOL (Apr. 8, 2015), https://law.yale.edu/admissions/profiles-statistics/student-perspectives/yale-federalist-society-lawyering-right.

15. ANN SOUTHWORTH, LAWYERS OF THE RIGHT: PROFESSIONALIZING THE CONSERVATIVE COALITION 116 (2008) (describing the experiences of conservative lawyers who said that their professional identities had been affected through hostile conflicts with liberal and radical professors and classmates); Avrahm Berkowitz, *The Day Free Speech Died at Harvard Law School*, OBSERVER (Apr. 3, 2016, 8:30 AM), http://observer.com/2016/04/the-day-free-speech-died-at-harvard-law-school/; Lisa Marie Passarella & David S. Jonas, *Political Disequilibrium on Law School Campuses*, ABA: BEFORE THE BAR (Jan. 12, 2017), http://abaforlawstudents.com/2017/01/12/political-disequilibrium-law-school-campuses/; Kelly Wilz, *The Myth of the Liberal Campus*, HUFFINGTON POST (Feb. 27, 2017) http://www.huffingtonpost.com/entry/the-myth-of-the-liberal-campus_us_58b1bc00e4b02f3f81e44812 (arguing that on many campuses liberal faculty feel silenced and unable to express views that might in any way be seen as political).

16. Kelly Wilz, *supra* note 15.

17. Robert D. Putnam, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 277-84, 340-41 (2001); Richard H. Pildes, *Why the Center Does Not Hold: The Causes of Hyperpolarized Democracy in America*, 99 CALIF. L. REV. 273, 275-81 (2011); Jonathan Haidt & Sam Abrams, *The Top 10 Reasons American Politics are so Broken*, WASH. POST (Jan. 7, 2015), https://www.washingtonpost.com/news/wonk/wp/2015/01/07/the-top-10-reasons-american-politics-are-worse-than-ever/?utm_term=.a26665c91a72.

18. Rosa Inocencio Smith, *Will Trump Voters and Clinton Voters Ever Relate?*, (Nov. 30, 2016, 12:58 PM), THE ATLANTIC, https://www.theatlantic.com/notes/2016/11/thanksgiving-politics/509165/.

two of the most commonly used but ineffective long-term approaches to handling and resolving conflict.[19]

III. CREATING SPACES FOR POLITICAL DIALOGUE: OPPORTUNITIES AND CHALLENGES

Over the past few years my colleagues and I at the Harvard Negotiation & Mediation Clinical Program ("HNMCP") have attempted to address the challenge of political polarization and complicit silencing in the classroom (and other spaces) by working with students to design and host dozens of facilitated political dialogue sessions.[20] In these sessions, trained facilitators invite those who hold different or opposing political views to participate in conversations on polarizing topics.

The primary purpose of these facilitated sessions has been to create a low-risk space where those with opposing views could simply feel comfortable enough to express their own views, have those views acknowledged by the others present in the group, and then be ready to listen to the views and reasoning of peers who might hold different or opposing views. Problem-solving, identifying common ground, writing a joint press release, or other "next steps" were not part of these sessions.

Only after hosting dozens of dialogues since 2013 was I able to observe—and then offer a name for—two other related (but perhaps even more troubling) dynamics that seem to be contributing to the political polarization of the moment:

(1) There exists a seeming lack of appreciation or respect for the value of simply sitting in the presence of openly-expressed differences and conflict, especially in situations where "problem-solving"—or some other joint "coming together" or "moving forward" activity—is not part of the agenda;

(2) Second, quite apart from the lack of appreciation for the value of "sitting with" conflict was the surprisingly low level of interpersonal skill and the dearth of practical tools that many individuals had for engaging others constructively around our political differences in face-to-face exchange. Indeed, observing the way in which highly talented, thoughtful, and passionate participants engaged so clumsily around a conversation about their political beliefs helped explain at least part of the reason why avoidance seems so attractive.

It is in addressing these two particular challenges—awareness of the value of 'sitting with' conflict and ability/skill to actually 'sit with' conflict skillfully—that I believe conflict management practitioners and scholars may find their greatest opportunity to make a difference at this moment of political polarization.

---

19. *Understanding the Stress Response*, HARVARD HEALTH PUBLISHING (Mar. 18, 2016), https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response (identifying the fight-or-flight stress response).

20. *See, e.g., Harvard Community Dialogues*, HARVARD NEGOTIATION & MEDIATION CLINICAL PROGRAM (Apr. 21, 2017), http://hnmcp.law.harvard.edu/newsletters/harvard-community-dialogues/.

A. Promoting the Idea of Dialogue Disconnected from Problem-Solving: Building Conflict Resilience

First, with respect to low appreciation among fellow citizens of the value of being in the presence of conflict around political differences detached from problem-solving: When my colleagues and I first began organizing political dialogue sessions, we often received questions from potential participants that asked us what action plan, joint project, takeaway, or "next steps" would emerge from the conversation. When I suggested that the dialogue was neither intended to persuade anyone nor to identify joint work or next steps but simply for the purpose of learning other's perspectives beyond one's own, asking curious questions, and sharing one's own perspective, participants frequently seemed confused and even disappointed.

At times, I was asked, "What would be the point of such a session?" In at least a few instances, individuals who had signed up to participate cancelled after learning that the session was not intended to be a "debate" or to improve their "oral advocacy skills" or to result in an "action plan" but merely for the purpose of learning and dialogue.

On the one hand, I am sympathetic to those who are skeptical about the idea of spending precious time in a room with others with whom they disagree and where the desired end-product is nothing more than the possibility of increased understanding (perhaps, if things go well!) and a sense that your own opinion has been registered with someone on the other side of the political fence. What value could there be in sitting in a room with those who have strongly-held and opposing political views from one's own when there is no intention to identify a solution or a way-forward? In a world where we rarely have enough time to be with the people we want to be with, what value could there be to sit awkwardly and uncomfortably with people who, at best, might make us uncomfortable and, at worst, might hold views that, if enacted as public policy, would inflict genuine damage on us or on those whom we love?

During the past two decades, law schools—as well as other professional schools—have re-oriented their curricula around the relentless and single-minded goal of "problem-solving."[21] As I read about current efforts even by those of us in the conflict management field to bring people with varying viewpoints together for dialogue, so many of these efforts seem focused on problem-solving as a main or primary goal.[22] To be clear, I am an advocate of problem-solving; indeed, much of

---

21. Joseph William Singer & Todd D. Rakoff, *Problem Solving for First-Year Law Students*, 7 ELON L. REV. 413 (2015); Katherine R. Kruse, Bobbi McAdoo & Sharon Press, *Client Problem Solving: Where ADR and Lawyering Skills Meet*, 7 Elon L. Rev. 225 (2015); Bobbi McAdoo, Sharon Press & Chelsea Griffin, *It's Time to Get It Right: Problem-Solving in the First-Year Curriculum*, 39 WASH. U. J.L. & POL'Y 39 (2012); John Lande & Jean R. Sternlight, *The Potential Contribution of ADR to an Integrated Curriculum: Preparing Law Students for Real World Lawyering*, 25 OHIO ST. J. ON DISP. RESOL. 247 (2010).

22. *See, e.g.*, The Divided Community Project, *Key Considerations for Community Leaders Facing Civil Unrest: Effective Problem-Solving Strategies that have been used in Other Communities* 3, OHIO ST. UNIV. MORITZ COLLEGE OF LAW (Jan. 2016), http://moritzlaw.osu.edu/dividedcommunityproject/wp-content/uploads/sites/101/2016/05/Key-Considerations-January-2016.pdf (outlining an approach to dialogue that is linked to community problem-solving); *Civic Engagement Facilitation Training: Moving from Dialogue to Collaborative Action*, ESSENTIAL PARTNERS, https://www.whatisessential.org/news/civic-engagement-facilitator-training-moving-dialogue-collaborative-action (Sept. 6, 2017) (offering a training designed to use dialogue for the purpose of collaborative problem-solving).

my teaching for the past two decades has been to help individuals and groups improve their individual and joint problem-solving skills.

At the same time, I worry that this relentless focus on "problem-solving" has diminished and obscured the independently worthy goal of simply building "conflict resilience," a skill that is essential in a pluralistic and diverse society where not every difference can be or will be resolved, but where we still need to find ways to coexist and work with each other peaceably and constructively in order to thrive and survive. I am not alone in this worry. Bernie Mayer has written extensively about the essential skill of engaging deeply with enduring conflict and the role that conflict specialists can and ought to play in helping individuals and organizations learn to sit with conflict.[23]

"Conflict resilience" is the ability to sit with and be fully present around those with whom we have fundamentally different views on critical issues.[24] Conflict resilience matters. When we sit in the presence of others with whom we may disagree strongly but with whom we can maintain civility and curiosity, we inevitably discover domains of shared interest and connection. And, even when we do not find these, we can often develop an appreciation for why our fellow citizens may hold the views they do. This "sitting with" does not solve an immediate problem; but it prevents the kind of demonization and othering that can escalate and cause new problems down the road while promoting humanization and connection. Sitting with conflict is not taming conflict or pretending it is not there. It is discussing it openly, acknowledging the tension and challenge it can create, but then refusing to let the different views become a corrosive force that blinds us to the reality of common humanity.

When I think about one of the most important and unique contributions that those of us in the conflict management field can offer at this moment of global political polarization—when technology is sorting us by political preference, income level, buying behavior, and consumer preferences—being a forceful voice for "conflict resilience," for the value of just sitting with and in the presence of those with whom we have strongly divergent views, may be among the most important contributions we can make. At times, this will mean cultivating patience in ourselves and opening ourselves up to some criticism from 'pragmatists' or 'problem-solvers' for being too "touchy-feely" or for not responding urgently enough to the exigency of the moment, or for not generating real "solutions" quickly enough.[25] Despite this, the conflict management field can stand as a witness to the value and the power of being present in the face of uncomfortable conflict and to the ways that this presence humanizes, decreases demonization, and identifies domains for cooperation even if it does not generate a "solution" to the presenting substantive conflict.

---

23. *See generally* BERNARD MAYER, STAYING WITH CONFLICT: A STRATEGIC APPROACH TO ONGOING DISPUTES (2009).

24. DIANE MUSHO HAMILTON, EVERYTHING IS WORKABLE: A ZEN APPROACH TO CONFLICT RESOLUTION 5, 91-92 (2013); MAYER, *supra* note 23, at 207, 235.

25. *See, e.g.,* MAYER, *supra* note 23, at 247 (discussing the challenge of insisting on a 'settlement orientation' that can encourage a rights-based approach to the resolution of conflict); ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: THE TRANSFORMATIVE APPROACH TO CONFLICT (2nd ed. 2004).

Technology has made it easier for us to escape our differences.[26] We live in curated worlds of interest through social media and we often find ourselves insulated from opposing viewpoints or those who make us feel unpleasant or uncomfortable—whether through choice or by dint of a Facebook algorithm. But because it is unlikely, and I would argue undesirable, for humans to ever simply be on the "same page" about all important issues and because it is essential in an ever-more connected and global world that we find ways of "being with" each other nonetheless in order to survive, developing an appreciation of the value of conflict resilience seems critical to me.

There is a fundamental and essential value in learning to sit and be present in the face of political differences and conflict. The pre-condition of problem-solving as an agenda item or purpose for the conversation with political nemeses can often get in the way of this worthy, independent goal.

B. Promoting the Skills of Conflict Resilience

Persuading individuals that "conflict resilience" is a skill worth developing is a challenging hurdle. Once it is achieved, however, the second area where conflict management professionals and scholars can contribute is in helping individuals build the skills necessary for "conflict resilience." Without a belief in the inherent value of "conflict resilience," though, it will be hard to persuade our fellow citizens or our students that they should take the time to develop the skills required for effective, meaningful civil discourse around deeply held political differences.

My experience training law students and observing facilitated dialogues around politically polarizing topics is that developing skills for facilitating and for participating in hard political conversations is incredibly challenging.

Let me start with challenges that many facilitators face: In their well-meaning attempt to create a low-risk space for dialogue participants, facilitators may be apt to 'over-design' conversations in ways that stress commonalities between the parties instead of bringing differences to light. Facilitators can be apt to design sessions that spend too much time on rapport-building or on activities that only touch gingerly and tangentially around the edges of truly hotwire political issues. In so doing, however, the fundamental differences between the participants that go to the actual heart of the political divide often remain unspoken, muted, or quickly extinguished and patrolled if inadvertently expressed by dialogue participant.

The seeming aversion by many facilitators to promote open expression of strong political differences is understandable: Unlike even a generation or two ago, there currently exist relatively few examples of healthy, robust, open, and respectful dialogue across political differences in our culture.[27] Instead, in their place, are plentiful demonstrations of less-than-edifying, shrill, and cheap political diatribes-masking-as-debates. Wanting to avoid the latter at all costs and not knowing what

---

26. Walter Quattrociocchi, Antonio Scala & Cass R. Sunstein, *Echo Chambers on Facebook* (June 13, 2016), https://ssrn.com/abstract=2795110.

27. SUSAN HERBST, RUDE DEMOCRACY: CIVILITY AND INCIVILITY IN AMERICAN POLITICS 133 (2010); Ray Williams, *The Rise of Incivility and What To Do About It*, PSYCHOL. TODAY (Oct. 21, 2016), https://www.psychologytoday.com/blog/wired-success/201610/the-rise-incivility-and-what-do-about-it.

the former might look like, facilitators unwittingly conspire with participants to ensure conflict stays out of the room.[28]

In addition, finding a way into deep and genuine dialogue around political differences takes time. It can be challenging for a group to feel comfortable enough to engage at a deep and personal level during a 90-minute or two-hour dialogue. And so, in many cases, they simply do not. It is for this reason that some of the most successful dialogue programs are sustained and iterative.[29]

While the avoidance-by-design approach to facilitation of hard political dialogue is understandable, it is also problematic for the missed opportunity it presents. As conflict management professionals, we need to do better at encouraging ourselves—and the people with whom we work—to lean in on differences and conflicts more directly. Yes, it may get emotional; yes, someone may feel a bit shaken by something said by someone else in the room; but that is part of what makes the work of deep engagement real, sustaining, and true; and that is part of what it means to live in a pluralistic and diverse society. We should not shun from this in a dialogue session; we should encourage it and do what we can to design the space to make it possible for such exchanges to occur.

Of course, it is not just facilitators who are often less-skillful in leaning into conflict and encouraging expression of differences. Dialogue participants often conspire—sometimes consciously, sometimes less-so—with facilitators in avoiding engagement around the toughest areas of difference in a dialogue.[30] Sharing one's views when one knows that others in the room may form negative impressions of them takes a high degree of tenacity, maturity, and vulnerability. Short of being pushed by a facilitator, participants in political dialogue may find it hard to muster the energy and courage to engage at the deepest level.

Participants in dialogue frequently report to me that they are of two minds when they are participating in a facilitated session: On the one hand, their agreement to participate in the conversation itself is motivated precisely because they *want* the chance to interact with those who have opposing views from their own. At the same time, at a conscious (and sometimes subconscious level), personal sharing around these issues feels very risky for them. Once these courageous individuals are in the room, it can be easier to simply collude with a facilitator who has designed a session around seemingly endless (and possibly even fun) introductory exercises, ground rules, and arms-length sharing at the edges of the polarizing issues rather than "getting real" and going deep. Indeed, one of the fears of going deep for many participants is the realization that they will realize there is no "solution" to the "problem." Fearful of directly confronting this painful reality in the room makes avoidance an even more attractive option. Yet, in my experience, participants who engage in conversations that fail to get to the heart of a particular issue, nonetheless often report regret and disappointment afterward; they may be less likely to return to a second session because of their experience failed to meet their expectations.

---

28. MAYER, *supra* note 23, at 68-80.

29. Harold H. Saunders et al., *Framework for Analysis: Sustained Dialogue, in* SUSTAINED DIALOGUE IN CONFLICTS: TRANSFORMATION AND CHANGE 23, 23-30 (Harold H. Saunders ed., 2011); Haggai Kupermintz & Gavriel Salomon, *Lessons to Be Learned From Research on Peace Education in the Context of Intractable Conflict*, 4 THEORY INTO PRAC. 293, 294-95 (2005).

30. MAYER, *supra* note 23, at 56-60 (citing nine major reasons why those in conflict often avoid direct engagement: fear, hopelessness, uncertainty, energy conservation, systems or relationship preservation, powerlessness, shame or embarrassment, inadequate skills, and resource depletion).

At least part of what holds participants back in dialogue sessions—in addition to some of the issues around social media and reputation that I already discussed—is a lack of skill in articulating their viewpoints in ways that are genuine, that can be heard by the other side, and that nonetheless avoid sweeping statements, attribution, and blame.

Conflict management professionals are uniquely placed to assist in equipping participants in dialogue with the skills to speak truthfully and powerfully, but also in ways that can be heard by others.[31] At their best, conflict management professionals can also design dialogue spaces where people are given permission to be less-than-articulate when they are struggling to communicate but are unable to find quite the right words. For the past six years I have worked with Seeds of Peace ("Seeds"), a non-profit organization that convenes and promotes dialogue work in several conflict zones, including the Israeli-Palestinian context, one of the most enduring and politically polarized conflicts in the world. I have observed many facilitators at Seeds articulate a concept they call "being raggedy." When a dialogue participant has something they want to express but is not sure how to phrase it articulately, rather than just remaining quiet, the participant asks the group for "permission to be raggedy." This is a signal to others in the session that what they are about to say may not be as well-formed as they prefer. Receiving permission to be "raggedy" encourages others in the group to accord an extra measure of grace and good intention before making judgments about the speaker. "Permission to be raggedy" invites the others in the room to put on their "curiosity hats" instead of their debate or judgement ones.

IV. CONCLUSION

As conflict management professionals think about what role we might play in a world of increased political polarization, I would urge our field to de-emphasize the problem-solving components of what we do and to amplify the independent value of engaging with those who hold deeply held political differences. Building "conflict resilience" skills and learning to sit in the midst of open conflict helps to humanize the other and creates the kind of environment that might lend itself to problem-solving in areas where possible and to respectful co-existence and increased empathy and understanding in those where it is not. If nothing else, conflict resilience avoids the creation of new problems—namely villainizing and dehumanizing the other—at the same time it provides opportunities to avoid conflict escalation by dint of bottling up hard feelings and strong views that otherwise boil over and can lead to destructive violence.

---

31. E. FRANKLIN DUKES, RESOLVING PUBLIC CONFLICT: TRANSFORMING COMMUNITY AND GOVERNANCE 69-71 (1996).

Hofstra Law Faculty Scholarship

2018

Living with No: Political Polarization and Transformative Dialogue

Erik Cleven

Robert A. Baruch Bush
Maurice A. Deane School of Law at Hofstra University

Judith A. Saul

Recommended Citation

Living with No: Political Polarization and Transformative Dialogue

Erik Cleven*, Robert A. Baruch Bush**, and Judith A. Saul***

I. INTRODUCTION

Political polarization is a fact in the United States and has been for some time.[1] Voters are more ideological and political parties are less open to compromise and bipartisanship. This has led to government shutdowns and legislative stalemate in Washington, and increasing difficulty for people to have productive conversations about politics. Political communication on television and online is less civil than face to face communication.[2] In many cases people withdraw from political participation as a result of polarization, exercise self-censorship and experience stress and even trauma. Political polarization also narrows individuals' understanding of themselves and their relationships with others. It makes it difficult to live with difference and diversity.[3]

This gives rise to the question of whether dispute resolution processes can help address the challenges political polarization raise. We argue that dispute resolution processes should not be seen as a substitute for the political process, but rather a complement that can help strengthen it. Based on this view, and on the authors' experience with dialogue work in the former Yugoslavia, as well as in urban and rural settings in the United States, we argue that transformative processes, specifically an approach we call *Transformative Dialogue*, are best suited to addressing the challenges of political polarization both in the United States and internationally. This is because the primary goal of transformative processes is not to reach agreement or find common ground, but rather to change the quality of conflict interactions from negative and destructive to positive and constructive. Transformative dialogue is about helping people gain their voice and choose identities and interactions that otherwise would be closed to them. It does so by supporting participants

---

* Erik Cleven, PhD, is Associate Professor, Department of Politics, Saint Anselm College, Manchester, NH; Board Member, Institute for the Study of Conflict Transformation (ISCT).
** Robert A. Baruch Bush, is Rains Distinguished Professor of Law, Maurice A. Dean School of Law, Hofstra University, Hempstead, NY; Founding Fellow and Board Member, ISCT.
*** Judith A. Saul, is Former Director, Community Dispute Resolution Center, Ithaca, NY; Certified Transformative Mediator; Fellow and Board Member, ISCT.

1. Many, but not all scholars would support the claim that America is politically polarized. For an alternative view *see* Morris P. Fiorina, Samuel A. Abrams, and Jeremy C. Pope, *Polarization in the American Public: Misconceptions and Misreading*, 70 J. POL. 556 (2008). For an overview of the debate between these scholars and critics, see *id.* at 556. Several of those claiming that America is polarized also make the point that this polarization has a long history, so if the question is whether polarization is new, there might be more agreement. *See, e.g.*, JAMES E. CAMPBELL, POLARIZED: MAKING SENSE OF A DIVIDED AMERICA (2016); ANDREW HARTMAN, A WAR FOR THE SOUL OF AMERICA: A HISTORY OF THE CULTURE WARS (2015).

2. DIANA C. MUTZ, IN-YOUR-FACE POLITICS: THE CONSEQUENCES OF UNCIVIL MEDIA 2 (2016).

3. In many other parts of the world, ethnic divisions also polarize politics, with conflict not just over political ideology, but even basic constitutional questions and therefore control of the state. DONALD L. HOROWITZ, ETHNIC GROUPS IN CONFLICT 187-88 (1985).

in gaining clarity about themselves and their interaction with others, and in considering the perspectives of others while maintaining their otherness.

II. POLITICAL POLARIZATION

The polarization of American politics is often identified by commentators as a serious challenge for American democracy.[4] In fact, politics in America have been polarized for some time. This is the case for voters, politicians, and political parties. There are significant differences among voters, with fewer moderates overall as well as fewer liberals in the Republican Party and fewer conservatives in the Democratic Party.[5] Ratings of Republican and Democratic politicians by interest groups like the NRA or Planned Parenthood overlap less and less over time.[6] These developments have taken place at the same time that membership in civic associations has declined, accompanied by a decrease in generalized trust,[7] while online communication is increasing. When people talk face to face they follow norms of politeness and civility; online communication not only facilitates uncivil political discourse, but gives those discourses a large audience.[8] The term "culture wars" also signals that views on cultural issues increasingly correlate with political identity.[9] This makes communication more challenging because there is less agreement about basic principles or outcomes. The other side becomes more "other," more alien, and understanding the other side is harder to imagine.[10]

Political polarization has serious consequences which challenge the functioning of democracies. In a polarized environment, people self-censor themselves and participate less because they fear the reactions of others who hold different opinions than their own.[11] The natural tendency of individuals to socialize with others like themselves is strengthened. While this can have positive effects, like promoting participation because of the in-group trust generated, it also lowers tolerance of others because individuals are not exposed to others with differing viewpoints.[12] These tendencies reinforce and lead to more polarization. After the 2016 elections many

---

4. Matthew Levendusky and Neil Malhotra, *The Media Makes Us Think We Are More Polarized Than We Really Are*, POLITICAL POLARIZATION IN AMERICAN POLITICS 106 (John Sides & Daniel J. Hopkins, eds., 2015).

5. CAMPBELL, *supra* note 1, at 2.

6. NOLAN M. MCCARTY, KEITH T. POOLE, & HOWARD ROSENTHAL, POLARIZED AMERICA: THE DANCE OF IDEOLOGY AND UNEQUAL RICHES 5 (2016).

7. *See* ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 49 (2000).

8. MUTZ, *supra* note 2.

9. HARTMAN, *supra* note 1, at 15.

10. In the United States this polarization exists against a backdrop of overall consensus on constitutional issues. Political conflict centers on hot issues like abortion, same-sex marriage, and the role of government in people's lives. In other countries conflict is about constitutional issues and the very nature of the state. In Kosovo, for example, Serbs are resisting inclusion into what is fundamentally an Albanian state. *See* Oison Tansey, *Kosovo: Independence and Tutelage*, J. DEMOCRACY 153, 154 (2009). Minorities in ethnically divided countries face the constant threat of a "tyranny of the majority". *See* ALEXIS DE TOCQUEVILLE, DEMOCRACY IN AMERICA 250 (1988). In some cases, fear of this tyranny can lead to violence and ethnic cleansing.

11. *See* Andrew F. Hayes, Dietram A. Scheufele, & Michael E. Huge, *Nonparticipation as Self-Censorship: Publicly Observable Political Activity In A Polarized Opinion Climate*, 28 POL. BEHAVIOR 259 (2006).

12. *See* DIANA CAROLE MUTZ, HEARING THE OTHER SIDE: DELIBERATIVE VERSES PARTICIPATORY DEMOCRACY 77 (2006).

people expressed strong emotional reactions including uneasiness, fear, and anger.[13] Indeed, CNN ran a headline using the term "post-election stress disorder."[14]

Political polarization and the state of American politics today clearly have serious consequences and there is reason to be concerned about how this will affect American democracy in the long run. Lowered participation and trust, self-censorship, the erosion of civility in political discourse and stalemate in Washington not only weaken the current functioning of democracy, but also ultimately can weaken citizens' very belief in democracy as a legitimate political system.

All this raises the question of whether dispute resolution processes can contribute to addressing these serious challenges. We argue that the right kind of dispute resolution process can do so. It must give people a voice and allow them to choose how to understand themselves and their relation to others and, especially, to live with difference. To do this, a dispute resolution process must not promise to "solve" problems or focus on "getting to yes." Rather it must allow people to disagree while still acknowledging the fundamental humanity of those on the other side. If the process can create more positive conflict interactions, people can disagree yet still live and work together. In short, these processes can help address political polarization if, rather than focusing on "getting to yes," they help us to learn to "live with no."[15]

We argue that transformative approaches to conflict,[16] particularly transformative dialogue,[17] are best suited to addressing these challenges because they do not seek to establish common ground but instead focus on the quality of conflict interaction, whatever the outcome of the process and the frequency or infrequency of future interaction.

---

13. A Pew Research Center survey found that while 51% of voters were hopeful after Trump's victory, 53% felt uneasy, 41% sad, and 31% angry. Pew Research Center, *Low Marks for Major Players in 2016 Election —Including the Winner*, Nov. 2016, at 2. When broken down by vote choice it is clear that among Clinton voters 91% felt uneasy (for Trump voters the figure is just 13%), 76% felt scared (5% for Trump voters) and only 7% felt hopeful (96% for Trump voters). *Id.* at 11. Clinton voters used words like shocked, disappointed and disgusted most often to describe how they felt after Trump's election with words like horrified, sad and devastated also frequently mentioned. *Id.* at 7. Nonetheless, 60% of Republicans expressed a wish for the GOP to be more conservative and 49% of Democrats expressed a wish for the Democratic Party to be more liberal, results suggesting that polarization may get worse, not better. *Id.* at 24. The authors' experience working with dialogue confirms this tendency toward differing perceptions of political reality and lack of desire to understand the other. When we have told some people that we are facilitating dialogue between liberals and conservatives, some people have expressed an aversion to the very idea of speaking to someone from "the other side." Some people express fatigue and exhaustion at having to constantly think about politics and race relations. However, we also find people longing for meaningful conversation and listening across the political divide. This attests to people's desire for moral connection to others. *See infra* notes 18-22, 38 and accompanying text.

14. Jenny Gold, *'Post-election stress disorder' strikes on both sides*, CNN (Feb. 20, 2017, 6:23 AM), http://www.cnn.com/2017/02/20/health/post-election-stress-partner/index.html.

15. ROGER FISHER, WILLIAM URY, & BRUCE PATTON, GETTING TO YES: NEGOTIATING AGREEMENT WITHOUT GIVING IN (1991). In this classic book, the authors argue that principled or "win-win" bargaining is the best approach to disputes of all kinds, and helps parties "get to yes". The authors of this article argue that in a polarized polity, "living with no" is a valid goal and important achievement.

16. *See* ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: THE TRANSFORMATIVE APPROACH TO CONFLICT (2nd Ed., 2005).

17. *See* ERIK CLEVEN, WHO NEEDS TO TALK TO WHOM ABOUT WHAT AND HOW?: TRANSFORMATIVE DIALOGUE IN SETTINGS OF ETHNOPOLITICAL CONFLICT (2011).

III. TRANSFORMATIVE DIALOGUE

A. The Transformative Approach to Conflict

The foundation for the transformative approach to conflict was laid with the publication of the first edition of *The Promise of Mediation* in 1994.[18] The book challenged the outcome-oriented nature of standard mediation processes and argued that even though many mediators tried to let solutions come from disputing parties themselves, the pressures to show the success of mediation, and the measurement of success primarily by the number of agreements reached, meant that in reality most mediation was outcome-oriented and mediator driven rather than party driven.[19] The transformative approach was based on several core premises that led to the unique nature of transformative practice.

The first of these premises is that human identity is defined by a balance between concern for self and concern for others. Human beings have a need for self-expression and independent agency. At the same time, individuals value connection to and understanding of others.[20] When conflicts occur this balance is disturbed and human identity is threatened; the transformative approach therefore claims that conflict is best understood as a crisis in human interaction.[21]

In addition to this, the transformative approach is based on the idea that people have the inherent *capacity* for self-determined choice and responsiveness to others.[22] However, the experience of conflict diminishes this capacity. Therefore, third party processes that focus on empowerment -- gaining more clarity about one's situation and the choices one faces – and recognition -- the ability to take the perspective of others -- are best suited to helping individuals regain the capacity for choice and responsiveness to others.

Operating from these premises means that transformative processes are genuinely party driven.[23] The key question an intervener first asks of the parties is, "[w]ho needs to talk to whom, about what, and how?"[24] This question acknowledges that people themselves are best positioned to decide who needs to be part of a conversation, what the conversation needs to be about and how they can best have that conversation. By contrast, most mediation approaches are outcome driven and

---

18. ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION: RESPONDING TO CONFLICT THROUGH EMPOWERMENT AND RECOGNITION (1994).

19. *See* Robert A. Baruch Bush, *Staying in Orbit or Breaking Free: The Relationship of Mediation to the Courts Over Four Decades*, 84 N.D. L. REV. 705, 727-32, 735-38 (2008); Robert A. Baruch Bush & Joseph P. Folger, *Mediation and Social Justice: Risks and Opportunities*, 27 OHIO STATE J. ON DISP. RESOL. 1, 22-28 (2012) (both articles referencing and summarizing research that documents the tendencies stated in the text).

20. BUSH & FOLGER, *supra* note 16, at 59-62 (arguing that this view of human identity is reflective of a "relational worldview" emerging in many fields and disciplines).

21. *Id.* at 45-62 (contrasting the transformative view with rights-based and problem-solving views of conflict).

22. *Id.* at 54-58.

23. *See* Bush and Folger, *supra* note 19, at 37-45 (describing the thoroughly party-driven character of transformative intervention practices).

24. *See* CLEVEN, *supra* note 17.

founded either on a human needs approach[25] or a narrative approach.[26] Processes based on both of these approaches tend to control and limit party interaction, even if they are intended to liberate participants from "zero-sum" biases and oppressive "dominant discourses."[27]

B. *Approaches to Multiparty Dialogue*

Most dialogue models claim to be less outcome driven than standard mediation and are defined as more open ended. For example, one definition of dialogue states that it is "a communication process that aims to build relationships between people as they share experiences, ideas, and information about a common concern."[28]

Saunders defines dialogue as

> ...a process of genuine *inter*action through which human beings listen to each other deeply enough to be changed by what they learn. Each makes a serious effort to take others' concerns into her or his own picture, even when disagreement persists. No participant gives up her or his identity, but each recognizes enough of the others' valid human claims that he or she will act differently toward the other.[29] [italics in original]

The problem with this definition is that it requires that participants listen to each other deeply and presumes that acting differently towards one another after the dialogue must be an outcome. In both of the definitions of dialogue cited above, the kind of speech allowed by the definition is limited. This requires that facilitators control the kind of speech taking place, and this can inhibit transformation because conflict is not fully expressed, and what is difficult is not confronted.[30]

---

25. *See, e.g.*, JOHN WEAR BURTON, CONFLICT: RESOLUTION AND PREVENTION (1993); FISHER, URY & PATTON, *supra* note 15.

26. *See, e.g.*, JOHN WINSLADE & GERALD MONK. NARRATIVE MEDIATION: A NEW APPROACH TO CONFLICT RESOLUTION (2001); Sara Cobb, *Empowerment and Mediation: A Narrative Perspective*, 9 NEG. J. 245 (1993).

27. In human needs approaches the third party translates what parties are saying into needs and interests language, which at best implicitly signals that the words the parties have chosen are inadequate or inappropriate, and at worst is patronizing, shutting people down unless they adopt the mediator's language. This is done in the interest of finding agreement. In narrative approaches, mediators "deconstruct" party narratives, only to reconstruct them in ways that the mediator has deemed is liberating and free of oppressive dominant discourses.

28. LISA SCHIRCH AND DAVID W. CAMPT, THE LITTLE BOOK OF DIALOGUE FOR DIFFICULT SUBJECTS: A PRACTICAL, HANDS-ON GUIDE 6 (2007).

29. HAROLD H. SAUNDERS, A PUBLIC PEACE PROCESS: SUSTAINED DIALOGUE TO TRANSFORM RACIAL AND ETHNIC CONFLICTS 82 (2001)._One reason that even Saunder's *"sustained dialogue"* is defined this way is that it, like most dialogue models, is based on ideas that resemble Martin Buber's idea of the I-Thou relationship. In an I-Thou relationship people do not see one another as separate, individual entities, but rather as a new whole. A more useful approach to dialogue could be based on the work of philosopher Emmanuel Levinas, who argues that the other person is radically Other and transcendent. But rather than seeing this difference as negative, Levinas shows that it is also the basis of freedom and moral choice. *See* EMMANUEL LEVINAS, TOTALITY AND INFINITY: AN ESSAY ON EXTERIORITY (1969) *[trans. Alphonso Lingis]*.

30. There is an argument to be made for limiting speech in dialogue. Mutz argues that cross-cutting interactions can lead to violence and that a certain level of civility or politeness is needed to reap the benefits of dialogue with those of differing views. *See* MUTZ, *supra* note 12, at 62. However, Transformative Dialogue is a facilitated process, where participants are not left to themselves but supported by a facilitator who makes it possible for strong confrontations to be potentially beneficial.

We suggest that dialogue can be successful even when no direct commu-
nication occurs – i.e., if third parties only speak to potential participants one-on-
one, or if dialogue only takes place with members of one group, i.e. *intra*group
dialogue, rather than between members of different groups (*inter*group dialogue).[31]
We suggest further that the role of dialogue facilitators is not to direct the kind of
interaction allowed, defining it from the start as listening to one another or acting
differently to one another, nor is their role to insist on particular forms of speech –
all of which are controlling, outcome-oriented practices. Rather, the intervener's
role is to support parties in reclaiming their capacity for moral choice, in deciding
on how to see themselves and on how and whether to recognize the other's perspec-
tive. This approach is what we call transformative dialogue.[32] We turn now to how
this can be done in practice and how it can address the challenges of political polar-
ization we outlined above.

C. *Transformative Dialogue*

Transformative dialogue is a process in which a third party works with different
members of a community in conflict to change the quality of interactions between
them in such a way as to increase the amount of pro-social interaction. The process
may include individuals as members of groups, social networks, organizations or
institutions. Pro-social interaction occurs when, regardless of how often or seldom
they interact, parties act from a position of clarity and strength and are open and
responsive to others whether they agree or have deep disagreements.[33]

Transformative dialogue is a multiparty process, like other dialogue ap-
proaches, and in multiparty processes identity plays a central and complex role.[34]
People participate both as individuals and as members of groups. But individuals
have many identities and these can be religious, ethnic, political, or social. A person
may be African-American, Jewish, Republican, and a father all at once. Each indi-
vidual has a unique understanding of their identity which is connected to their
unique history and life experience, but people also have elements of their identities

---

31. It is important to understand that the value of dialogue can be realized whether or not people meet
across group boundaries (intergroup dialogue). Even intragroup dialogue has value. No matter how
much we have a tendency to see groups as unified – the African-American community, the Jewish com-
munity, the Serb community – each of these groups consists of individuals and a plethora of views and
understandings. Often people cannot effectively come together for dialogue, at least as members of
groups, until those groups have had a chance to come together to consider how they wish to interact with
the other side. One of the authors was told this in no uncertain terms in Northern Kosovo by Serbs there.
Serbs were not ready for interethnic dialogue with Albanians because they hardly knew who they were
as a group and up until that point only the loudest, and therefore sometimes the most radical, had had a
chance to be part of the discussion. Meanwhile, conflict resolution intervenors were only inviting mem-
bers of the Serb community who agreed with the international agenda to seminars, in order to avoid
"spoilers" and so-called difficult people.

32. *See* CLEVEN, *supra* note 17.

33. The definition in the text is the authors' own usage based on transformative conflict theory. *See*
BUSH & FOLGER, *supra* note 16. Broader definitions are found, especially in the literature on social
psychology, *see, e.g.,* C. DANIEL BATSON & ADAM A. POWELL, *Altruism and Prosocial Behavior,*
HANDBOOK OF PSYCHOLOGY 463 (2003). However, the definition here is generally consistent with those
broader ones. The point of the text is not that a facilitator *directs* the parties toward pro-social interaction,
but that such interaction is the *result* of the process.

34. *See* AMIN MAALOUF, IN THE NAME OF IDENTITY: VIOLENCE AND THE NEED TO BELONG 12-15
(2003); CHARLES TILLY, IDENTITIES, BOUNDARIES AND SOCIAL TIES 8-9 (2005).

that are shared with other members of their groups.   Furthermore,   individuals' understanding of their own and others' identities changes as they interact.[35]

When people experience conflict, the transformative model suggests, they become weak and self-absorbed.[36] One important element of the weakness that occurs in conflict is the narrowing of identities: one particular part of who a person is may become most important, even though that person has many other identities. And the part of identity that is shared with other group members becomes all important, at the expense of the others.[37] Self-absorption at the group level is basically polarization, as was already explored in the discussion of political polarization above. People also feel that their group has been uniquely victimized. Sometimes this is of course the case, but often this ignores the reality that victims and aggressors exist on both sides of conflicts.

Dialogue facilitators in other approaches strive for "balance" – working to get similar numbers of people from pre-determined "sides" or groups to participate.[38] Facilitators of transformative dialogue also consider individuals' group membership. But rather than defining who the groups and individuals are in advance, a transformative facilitator will allow those categories and identities to emerge in party-driven fashion, through conversations with members of the community.[39] Our current transformative dialogue work in two communities in the United States provides good examples of the different ways people choose to identify themselves. One community is urban and the other rural.[40] In both, as discussed below, the first steps toward dialogue were initiated by community members themselves, who had heard about transformative dialogue practice, and reached out to transformative facilitators for support.

In the urban community, people themselves define communities in the city in terms of race and ethnicity, and in part according to religion. African-Americans, Hispanics, and various religious denominations define some of the most important groups that people said needed to be involved. This may be because in the urban environment, policing and the interaction between authorities and citizens, to give one example, plays out differently for racial minorities than others. In the rural setting we are working in, political identity is foremost. People contacted us primarily because of a concern about the effects of political polarization on their community. "Trump supporters," "conservatives," and "liberals" are the terms most

---

35. JAMES A. HOLSTEIN & JABER F. GUBRIUM, THE SELF WE LIVE BY: NARRATIVE IDENTITY IN A POSTMODERN WORLD (2000).

36. That is, they lose the clarity they need to make decisions about what they want to do and how they want to relate to others, and they lose the capacity to connect constructively with others and see their perspective. *See* BUSH & FOLGER, *supra* note 16, at 49-53.

37. See MAALOUF, *supra* note 34.

38. *See, e.g.,* Margaret Herzig, *Moving from polarized polemic to constructive conversation–A report from the Public Conversations Project,* 7 J. PUB. PARTICIPATION 1 (2001) (describing the "structured dialogue" process used by one prominent organization, including preplanning of invitations and other "structured" elements described in the text below, *see infra* notes 39-42 and accompanying text); SCHIRCH & CAMPT, *supra* note 28.

39. *See* CLEVEN, *supra* note 17, at 13.

40. The dialogues in these two communities are current and ongoing, and the participants prefer that their work not be cited for publication by name or location, for privacy reasons. Other recent examples of transformative dialogue facilitation include a multi-session dialogue on "passive racism" between white and black residents in a southern Maryland community. *See* Richard "Dusty" Rhoades, *Journey into Self and Other, in* TRANSFORMING CONFLICT FROM THE INSIDE OUT 23 (Robert A. Baruch Bush & Joseph P. Folger eds., 2016).

often used to describe different people in this community in spite of the fact that, like their urban counterparts, they also differ in terms of ethnicity and religion. In both urban and rural engagements, the participants self-identified the groups that needed to be involved.

Another practice of transformative dialogue is that the facilitator does not push a particular content or goal for the dialogue. The facilitator does not come in with an agenda to engage participants about a pre-determined topic, but trusts that people know what they want to talk about and what they want to achieve; they also know when they want to change the topic of discussion for some reason. In the rural community we are working in, the concern with political polarization already determines that some topics are of more interest to people than others. In the urban community where we are working, participants requested a meeting several days after the August 2017, Charlottesville incident and its aftermath.[41] Though racism and white nationalism had not been the main focus of earlier sessions, the participants decided to change the topic, the facilitators responded quickly, and people were able to talk about the event's impact on them as individuals and on their community. By contrast, most dialogue processes involve the facilitator in shaping the agenda, either in advance or as the process unfolds, or both. This is another example of the way in which transformative dialogue is genuinely party-driven.

Finally, while most dialogue models give the facilitator the responsibility to set and enforce rules that are supposed to ensure civil exchange, transformative dialogue leaves responsibility for this to the participants themselves – so that heated exchanges and conflict within the dialogue are possible. From such unconstrained exchanges, participants can make lasting changes in their views of themselves and each other.[42] They are free to express their own "truth" and hear others doing the same. In a facilitated face to face exchange individuals have a chance to hear themselves speaking out loud. This often leads people to rethink what they are saying and how they are saying it, and when people do this on their own it leads to real change. Dialogues that have strict rules about how people interact may promote civility while the facilitator is enforcing the rules, but may have little or no long-term effect on interactions. In summary, allowing conversations to go to the heart of the participants' differences allows them to confront what is difficult, to take responsibility themselves for the exchange, and to hear each other's voices in a new way. All of these factors help contribute to real transformation, not simply the polite suppression of difference controlled by a third party.

D. How Transformative Dialogue Can Address the Challenges of Political Polarization

As mentioned earlier, dialogue in a democracy is not a substitute for the political process, nor should it be. In a democratic country, the political process can be

---

41. *See* Aaron C. Davis, Joe Heim and Laura Vozella, *How Charlottesville Lost Control,* WASHINGTON POST (Aug.26, 2017), https://www.washingtonpost.com/investigations/how-charlottesville-lost-control-amid-deadly-protest/2017/08/26/288ffd4a-88f7-11e7-a94f-3139abce39f5_story.html?utm_term=.e932283554d6.

42. *See* Judith A. Saul & Scott Sears, *A Relational Perspective on Multi-Party Practice, in* TRANSFORMATIVE MEDIATION: A SOURCEBOOK 397, 407 (Joseph Folger, Robert Bush & Dorothy Della Noce, eds., 2010) (arguing that "Conflict is a prevalent and potentially constructive element of group interaction.").

seen as a form of conflict resolution.[43] Many different opinions and preferences exist with regard to public policy questions, and a country's constitution sets out rules by which people deliberate, debate, advocate and lobby, and finally vote for a particular political representative or policy.

Dispute resolution processes, in order to effectively address the challenges of political polarization, must not supplant this fundamental process. However, they can support and strengthen the political process so that it becomes more constructive in the face of divisions and conflict, even when these conflicts seem intractable.[44] This is true whether or not people reach agreement or reconcile. Transformative processes are uniquely situated to contribute in this way because they do not push people to seek common ground. When conflicts are complex and closely related to identity, a more appropriate goal is to support people in finding ways to interact and relate in spite of their deep disagreements.

There is therefore value and impact in dialogue even when common ground is not sought – in fact we might even say *especially* where common ground is not sought. Transformative facilitators do not define people or their identities in advance. Instead, they follow Levinas in respecting people's radical difference.[45] They allow people to explore their identities, individually or as members or groups. The first main impact of transformative dialogue is that it allows people to choose the identity they wish to emphasize. However, this happens not in an isolated room, but in the real context of that person's community. Through dialogue, people gain clarity and strength to decide how they want to understand and live out their identity. And by interacting with others face to face, the complexity and multi-faceted nature of identity is shaped and constructed more fully.

For example, in one of the community dialogues mentioned above a participant remarked that after talking to a member of the sheriff's department directly, she changed her view of law enforcement officers. As the participant put it, "getting to know someone can change assumptions." Another person remarked that through the dialogue they realized they were more judgmental and angry than they thought they were. These examples attest to the greater clarity in understanding of self and other that emerges in transformative dialogue.

The second potential impact of transformative dialogue is that based on the strength and clarity gained through the process, people can make better decisions going forward as individuals and/or as members of groups. Hearing from others in their own group or from those in other groups often reveals that seemingly simple situations are more complex. And this complexity often opens possibilities that had seemed closed earlier. Sometimes participants may decide not to move forward with certain options because they realize the time is not right. This too is a positive result of a dialogue. Moreover, as mentioned above, a transformative dialogue usually starts through having conversations with individuals as well as small groups to determine who needs to talk to whom, about what, and how.[46] This naturally leads to a focus on local networks. Working with and through local networks rather than bringing individuals together in a pre-determined or random fashion increases the

---

43. *See* Lon L. Fuller, *The Forms and Limits of Adjudication*, 92 HARV. L. REV. 353, 363-65 (1978).

44. *See* Heidi Burgess & Guy Burgess, *Constructive Confrontation: A Transformative Approach to Intractable Conflicts*, 13 MEDIATION Q. 305 (1996).

45. *See* LEVINAS, *supra* note 29 and accompanying text.

46. *See supra* note 24 and accompanying text; CLEVEN, *supra* note 17.

possibility of post-dialogue impact, since people are already connected in ways that will continue.

The third potential of transformative dialogue is that it gives people the freedom to recognize the humanity of others.[47] By meeting the other face to face and seeing that they too are not just members of the other group, but unique individuals, and hearing their unique voice and perhaps the voices of other members of their own group in the room, it may be easier to recognize the other's humanity, all the while also recognizing that this does not require agreement or sameness.[48] A dialogue participant can choose to remain "civilly distant" from another.[49] That is still a large step forward from antipathy and hate.[50]

IV. CONCLUSION: TRANSFORMATIVE DIALOGUE AND POLITICAL POLARIZATION

Based on the above discussion, we argue that transformative dialogue can address the challenges of political polarization. First, it does not require common ground. Instead it recognizes that differences exist and that it is all right to "live with no." This makes dialogue more attractive to many because they do not feel threatened by it but see dialogue as a way to establish more constructive interactions across boundaries of definite difference. However, at the same time, the increased clarity and willingness to consider the perspective of others that are hallmarks of this process may sometimes lead to participants finding common ground.

Second, transformative dialogue can alleviate stress and fear because it contributes to recognition of others and their perspectives. As people gain more clarity through the dialogue process, and get to know the other, they also develop strength and confidence. Processes that drive participants towards reconciliation or that emphasize commonalities do not achieve this as effectively, because they are forced, and they often require participants to suppress the way they really want to talk about divisions. Instead, transformative processes confront what is difficult head on and help people deal with it.

Third, transformative dialogue supports people both in finding their voice and what they want to say, and also in being clear about what they do not want to say. The latter is very different from the kind of self-censorship described in the literature on political polarization.[51] If people are self-censoring themselves then they are holding back when they really have things they want to say. Choosing not to speak, from a position of clarity or strength, is different. Still, it may be that in a

---

47. We recognize that other approaches are also aimed at participants recognizing the humanity of others. Our point is simply that controlling the speech that is allowed in a dialogue runs counter to that goal.

48. In fact, as Levinas points out, acknowledgement of this otherness is also an acknowledgement of the freedom we have as human beings.

49. *See* CLEVEN, *supra* note 17, at 10-11 (citing this example: "When interethnic riots broke out, a man [who] was a local nationalist leader among his ethnic group ... went out in front of a group of his own people and persuaded them not to commit acts of violence against the other group. He later stated that not only would he not have done that had he not participated in the dialogue process, he would also not even have thought of it. He was not an advocate for reconciliation, but his relationship to members of his own group and members of the other ethnic group had changed [even though] he still had conflict with members of the other group...").

50. *See* AARON T. BECK, PRISONERS OF HATE: THE COGNITIVE BASIS OF ANGER, HOSTILITY AND VIOLENCE (2000) (describing the alienation typical of destructive conflict as a prison).

51. *See supra* text accompanying notes 10-12.

face to face dialogue people find the words they *do* want to say, because as the process progresses it moves from negative and destructive to positive and constructive.

For all of these reasons, discourse has the potential to become more civil in a transformative dialogue and to remain more civil outside the context of the dialogue. Because it allows people to move at their own pace, to speak honestly and in their own words, the changes that occur during a transformative dialogue are more likely to continue when people interact in the future.[52]

Political polarization is a reality in the United States and elsewhere in the world; it has serious consequences that can threaten the way democracy works and therefore also its very legitimacy. The transformative dialogue process can address these challenges. It can help people deal with issues that matter to them and support more constructive conflict interactions, regardless of whether or not people find common ground. In this way, it can create the basis for democratic deliberation within and across group boundaries that can lead to greater tolerance and a better functioning political process – so that even when we cannot get to yes, we can nevertheless live civilly with no.

---

52. *See supra* text accompanying notes 45-46.

Public Policy Mediation

By Howard S. Bellman and Susan L. Podziba

I ronically, there is no consensus on the definition of public policy mediation among those in the field. Nonetheless, it has been successfully applied to a great variety of policy challenges and belongs in the array of processes that promise to enhance public discourse. It is an outcome-oriented method for securing actionable agreements among identified parties, who participate as negotiators, often on behalf of constituencies. As such, it overlaps and complements civil court mediation as well as public engagement processes that promote and support civil discourse. These processes are important in helping us to resolve important public matters, such as Superfund cleanups, planning the 9/11 memorial, and sitting waste disposal facilities, in a civil manner.

Public policy mediation creates a forum for deliberative negotiations among government, stakeholders and, when appropriate, the public. The parties' contributions of their technical expertise and the resulting greater knowledge of participants' preferences, are woven into discussions that increase mutual understanding and lead to otherwise untapped opportunities for consensus agreements. The intended result is an agreement that sets forth the terms of the future relationships and responsibilities among the parties with regard to the issues they discussed.

With tailored applications across the policy spectrum — from agriculture to fishing, environment to nuclear power, education to health, and land use to transportation policy — public policy mediation appears in many forms. Of course, public policy mediators, like mediators in general, must be responsive to the unique characteristics of case situations. Typically, this involves attentiveness to layered complexities. The examples below — negotiated rulemaking; policy, planning, permitting, and development cases; and administrative prosecutions and court litigation — are provided to help illustrate its procedural and substantive scope.

Varieties of Public Policy Mediation

Negotiated rulemaking, which convenes government and stakeholders, has been used to formulate state and federal regulations on a broad range of issues including the protection of captive marine mammals, the administration of student loans, worker safety in the operation of construction cranes, and the licensing of radioactive waste disposal facilities.

In **planning, policy, permitting, and community development cases**, which require an interface among governmental entities, stakeholders, and citizens, public policy mediation has been used to reach agreements rooted in the diverse interests of the public and relevant parties. Examples include planning for the multiple uses of public forests; the protection of watersheds; the closure of public mental health facilities; and even the development of a city charter required for a municipality to emerge from state receivership authority.

Similarly, public policy mediation has been applied in matters initiated by private sector parties seeking governmental permission or public support for their proposals. Examples include proposed private developments that require approvals by planning and zoning authorities; permitting for waste disposal facilities; and forestry sector efforts to promote carbon emission trading programs.

Negotiated rulemakings and planning, policy, permitting, and community development cases, although usually circumscribed by statutory authorities and sometimes subject to court review, are typically not precipitated by litigation, or conducted "in the shadow of the court." Rather, they are initiatives of governmental authorities seeking broad participation in public decision-making to achieve fully informed, efficiently implementable outcomes. They are transactions necessary to move forward to a public objective.

On the other hand, **administrative prosecutions and court litigation** have precipitated public policy mediation in cases, for example, where river restoration terms were negotiable; Indian land claims and treaty rights were in dispute; school district racial integration was required; and for Superfund cleanups. In these matters as well, the settlements specify future relationships and responsibilities.

Defining Characteristics of Public Policy Mediation

Common factors that characterize public policy mediation include the necessity for governmental action of some kind; the capacity for "intervention" by numerous stakeholders, who are politically as well as legally relevant; and a potential for future "public" impact. It is also characterized by a reluctance of government to effectuate its interests unilaterally, or an expectation that

broad participation will increase its ability to efficiently achieve its policy goals. In addition, most of the negotiations, other than those based mainly on litigation or administrative enforcement procedures, are conducted under open meeting laws that allow the general public to attend and provide input throughout the process.

Procedurally, public policy mediation cases typically involve as parties multiple government authorities, advocacy and community nonprofit organizations, and private sector entities. The process provides a forum for such parties, who may represent thousands of constituents, to negotiate a set of complex issues and build an unprecedented consensus agreement that addresses a public issue. For example, among the dozen or more student loan negotiated rulemakings we've mediated, as many as 20 parties representing state and federal government agencies, all categories of higher education institutions, lenders and loan servicers, and students and borrowers, have developed consensus rules that impact all students who have received federal financial aid, their parents, and the schools they attend.

Public policy mediation includes a host of pre-negotiation activities, the first of which is an assessment. Policy mediators interview relevant parties to determine the agenda of issues, learn the history of past efforts to address the situation, and identify the dynamics likely to affect the negotiations. In addition, sometimes the assessment includes the need to specify or confirm the parties that must be involved to ensure any agreement is actionable, and/or to determine the feasibility of initiating negotiations. These early interviews also provide an opportunity for the mediators to build confidence in the mediation process and initiate relationships with negotiators.

In most cases, an initial negotiating task of developing organizational protocols serves to clarify expectations and to prevent procedural conflicts from arising during the process. These agreed upon terms specify many procedural understandings including the group's definition of consensus, the roles and responsibilities of the negotiators and mediators, the compliance to which ultimate agreement commits the parties, and the nature of appropriate media contact.

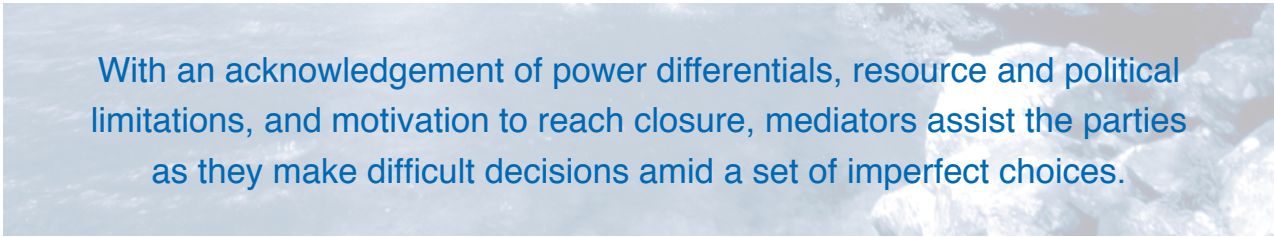Upon completion of such preliminaries, public policy mediators manage substantive issues to enable parties

Over the years since 1965, **Howard Bellman** *has mediated in nearly every category of dispute. His work has ranged from the most ordinary civil and labor matters to international diplomacy. A significant portion of his practice has included high-profile, multi-party cases of public concern such as controversial land-use determinations, large-scale environmental remediations, school district desegregation, state-wide education financing litigation, and Indian land claims.*

Susan Podziba *has served as a Public Policy Mediator for more than twenty years. Ms. Podziba served as Fulbright Senior Specialist in Peace and Conflict Resolution at the Amsterdam Centre for Conflict Studies at the University of Amsterdam, Netherlands and has taught graduate courses as a Lecturer at the Massachusetts Institute of Technology and as a Faculty Associate at the Program On Negotiation at Harvard Law School. Her forthcoming book,* Civic Fusion: Mediating Polarized Public Disputes, *will be published by the American Bar Association in the spring of 2012. For more information, go to http://www.podziba.com.*

Howard Bellman

Susan Podziba

to explore and deepen their understandings of policy issues of concern. The negotiators generate and refine options until they reach conceptual agreements. Such early agreements are then reflected in written drafts and repeatedly revised until consensus is achieved.

Ultimately, the participants' enhanced grasp of relevant factors, including the interests of others gained by carefully designed discussions, provides opportunities for agreements not previously perceived through the limited lens of each negotiator. With an acknowledgement of power differentials, resource and political limitations, and motivation to reach closure, mediators assist the parties as they make difficult decisions amid a set of imperfect choices.

Public Policy Mediation in Comparison to Other Processes

Public policy mediation is related to and overlaps substantially with civil litigation settlement mediation as well as with various public engagement processes. Among the key similarities and differences between public policy mediation and these other categories of process interventions are the objective of the process, the participating parties, and the process components and dynamics.

Civil Court Mediation

The Objective: The objective of both public policy mediation and civil court mediation is an actionable agreement. The latter typically results in settlements of disputes that would otherwise be adjudicated by a court, for example, negligence and breach of contract cases. Such settlements represent an application and enforcement of existing laws, and first and foremost, resolve disputes arising from past interactions. It may be said, however, that the evolution of civil court mediation has led to a growing appreciation of its value in "broad" disputes and the problem-solving capacity of transactional negotiations even in cases apparently precipitated by litigation. For example, the mediator who helps litigating parties develop a new business arrangement, in conjunction with settling a lawsuit over a past dispute, has added this relational element to the case evaluation he may have been expected to provide.

Public policy mediation outcomes are referred to as "settlements" for the subset of cases initiated in response to administrative prosecutions and litigation, and "consensus agreements" for those cases that result

in sub-statutory regulations, policies, plans, permits, or other forms of public agreements. In most instances, the agreement reached is a consensus recommendation to the government agency with jurisdiction over its implementation.

When the outcome is a consensus recommendation, organizational protocols agreed to prior to substantive negotiations often speak to the government's commitment to implementation. In some cases, for example, negotiated rulemakings, the consensus agreement is subjected to additional public review.

Most public policy mediation cases focus on building agreements related to future actions and activities. They are initiated because of existing or expected substantive and procedural disputes that may prevent the parties from addressing their shared issues. As a result, such mediated agreements establish the future relationships and responsibilities among the participating government entity and the parties and their constituents with regard to substantive issues. Public policy mediation, in that it creates arrangements to be applied in the future, is similar to mediation in collective bargaining, perhaps, historically the most common application of mediation.

The Participants: In civil court matters, the negotiating parties are usually predetermined as plaintiffs and defendants (or as divorcing spouses and their children, or unions and employers, as in family court matters and collective bargaining). In contrast, public policy mediation cases often require intensive pre-negotiation activities to identify representative negotiators, who may number in the dozens. For example, in negotiated rulemakings, an initiating agency may request stakeholder nominations through a formal federal register notice. After considering the nominees, it specifies negotiating committee members in a second notice and asks for public comment to ensure its participants represent a balanced group of interested parties. In other cases, as part of the assessment, the mediator may identify and recommend potential negotiators to represent stakeholders and/or segments of communities. The negotiators are the primary participants, and they must represent the interests of their constituents, and sometimes consider the opinions and preferences of the public at large.

Unlike civil court mediation and the administrative prosecutions and litigation subset of policy cases, most parties in public policy mediation cases are not

represented in the negotiations by legal counsel, although some may consult with counsel.

The Process: The process components of public policy mediation and civil court mediation have many similarities. Both rely on an assessment of the situation, protocols regarding how the process will proceed, and methods for managing negotiations.

Most mediators conduct some form of assessment prior to or early in the negotiations. In civil court cases, a diagnosis may be completed via a pre-negotiation exchange of documents or during early caucuses, or in complex cases, through pre-negotiation interviews. Before formal negotiating sessions, as indicated earlier, public policy mediators usually conduct intensive interviews with identified negotiators, and sometimes, with potential negotiators as well. Sometimes the findings of the assessment are documented and shared with the parties and the public.

Shared protocols provide process structure and clarify what is expected of the parties and the mediators. In civil court mediations, such protocols may be briefly stated or even implied by custom. Often, attorneys rely on the pre-existing norms of settlement negotiations.

In many public policy mediations, by contrast, the parties must develop organizational protocols to create the norms that will govern their negotiations and to make explicit their future commitments under agreements reached. Mediators assist the parties in formulating an agreed-upon protocol document prior to commencing substantive discussions.

Managing Negotiations: In all mediation cases, parties examine interests and develop options. There are caucuses and sidebars. The mediators work to avoid and overcome impasse and depend on the growing trust they develop with negotiators concerning confidence in the process and mediator competence. The role of "agents of reality" is familiar to all mediators.

The scope of issues typical to public policy mediation cases and the amount of time required to attend to them, is similar to the most complex civil court cases. In addition, since public policy cases typically involve comparatively large numbers of parties, negotiations require large group meeting facilitation as well as time and mechanisms for consultations with constituents.

Public Engagement Processes: Dialogue, Facilitated Meetings and Workshops, Facilitated Advisory Committees

Public policy mediation is also related to and overlaps with other processes that engage citizens in discussions of public concerns. Dialogues, facilitated meetings and workshops, and facilitated advisory committees are similar to and different from public policy mediation in relation to their intended objectives, the participants, and process components and dynamics.

Objectives: Public policy mediation and the above public engagement processes are generally initiated to provide forums for vigorous and open, yet safe, exchanges of ideas, opinions, and perspectives on public issues. The expected outcome of public policy mediation, as has been stated above, is a consensus agreement. Examples include a plan to reduce the incidental take of marine mammals during commercial fishing operations and the resolution of Indian land claims.

Dialogic processes generally focus on discerning similarities and differences, increasing understandings, mutual respect, and breaking down of stereotypes among people with differing world views and deeply held values. Examples include dialogues on the issue of abortion and inter-faith dialogues. Sometimes dialogue participants elect to take individual actions in light of their new understandings.

The results of *facilitated meetings and workshops* are usually recommendations rooted in participants' shared preferences and priorities concerning the issues under discussion. These recommendations, and the public support for them, are meant to inform and influence decision makers. Examples include meetings concerning siting a waste disposal facility, pandemic influenza planning, and the 9/11 redevelopment and memorial.

Facilitated advisory committees are typically convened to provide government officials with advice on a range of related issues and topics. Advice may be offered as consensus recommendations, in minority and majority reports, or as individual opinions. The government agency requesting advice may choose to adopt some or all of the committee's recommendations. Examples include a governor's task force on health care policy, federal advisory committee on environmental justice, and a citizen advisory group concerning a Superfund cleanup.

Participants: *Dialogues* typically involve individuals, who may act in their capacity as citizens or as representative of a group, for example, a religious leader; but they usually do not represent constituents. Individuals may voluntarily self-select for a particular dialogue, or a sponsor may identify and invite participants.

Facilitated meetings and workshops cover the broadest range of possible participants and may include citizens and/or technical experts and/or stakeholders or any combination thereof. As with dialogues, participants sometimes choose to attend or they may be invited to do so.

Citizens who choose to participate usually populate citizen *advisory committees*, as in a Superfund citizens advisory group. On the other hand, members of government advisory committees, such as those governed under the Federal Advisory Committee Act (FACA), are appointed by the sponsoring government agency to represent particular stakeholder groups or due to their technical expertise. Committees established under FACA must

include parties that represent the spectrum of interests and perspectives indicated by the issues under discussion.

In both government advisory committees and public policy mediation processes, the range of relevant interests are represented, and the public may be provided with opportunities to participate. However, in policy mediation the sponsoring government authority participates as an active negotiator, and the public involvement component may be very specifically designed to inform and influence the negotiators, as occurred in a consensus process to develop a city charter to restore local governance.

The Process: The process components of public policy mediation and other public engagement processes are similar in relation to required pre-meeting activities and differ in the structure and dynamics of the meetings. All require assessments to develop a meeting design specific to the substantive focus and goals; and in some cases, to identify or confirm the participants. Additionally, within the context of the processes, participants are furnished with trusted information, and they develop shared norms and expectations regarding how the meeting will proceed, its expected results, and how those results will be used.

All of these processes rely on facilitators to promote civil discourse on public issues, but with varied dynamics. *Dialogue* facilitators sustain participant focus on substantive issues as well as on how ideas and perspectives are conveyed. They work to surface underlying beliefs and assumptions in an effort to cause shifts in understandings and emphasize human connections despite differences in values and beliefs. *Facilitated public meetings and workshops* often involve facilitation teams, including a lead facilitator for plenary sessions and small group facilitators. In small groups, participants typically discuss the issues and generate ideas, all of which is reported and compiled. Priority recommendations and preferences of the large group may be identified as those common across multiple small groups or through a voting mechanism.

Advisory committee facilitators promote discussions that enable participants to pursue a great range of ideas and options. In these processes, the goal is to explore the issues and provide feedback and advice to the government. In some cases, it is not necessary to synthesize the feedback among the various participants as government officials attend as observers. In others, facilitators may assist the group in developing consensus or majority/minority recommendations on a topic discussed only at that meeting or on others that straddled multiple meetings. For example, at some meetings, an Iron and Steel Advisory Committee provided consensus recommendations for redeveloping closed industrial facilities and at other meetings, offered regulatory revisions. The initiating agency may accept of reject any recommendations.

In public policy mediation, mediators also facilitate discussions, however, with a narrower focus than other facilitated meetings because of the intent to prepare a written agreement. Analysis of options, for example, may be more limited due to existing laws, power differentials, and/or resource constraints that the agreement must respect. Additionally, after reaching conceptual agreements, participant discussion is focused on revising proposed written text into a consensus agreement. The process typically occurs over a period of months and includes mediator activity between meetings to promote efficiency at the plenary negotiating sessions.

It is important to understand that public policy mediation is not an alternative to any of these other public engagement processes. Indeed, whereas these processes may be designed to inform the public, interest groups, and government so as to mitigate and even avoid conflict in the development of public policy, mediation is often applied only after conflict is well developed. Dispute resolution experts understand that the array of processes constitute a repertoire of "process options" to be applied by an astute practitioner, who understands the capacities and limitations of each option. Moreover, these processes may be implemented in a sequence or in tandem so that, for example, facilitated meetings inform negotiators of public views and advise the public of the progress of negotiations.

The Future of Public Policy Mediation

Given the polarization and unusually contentious state of political behavior and strategies across the nation, we anticipate an increased reliance on public policy mediation in the near future. The general awareness of mediation continues to grow in the public and the media: recently editorials have called for mediation as public issues deadlock, and political leaders have appealed for mediation. It appears that there is an increasing appreciation of the potential value of mediation in the political process.

In well-publicized events in Wisconsin during the summer of 2011, for example, there have been requests for mediation to assist in the resolution of issues arising out of the occupation of the State Capitol by protesters; to support negotiations to bring home legislators, who had left the State to frustrate the passage of certain bills; and even to assist in the development of better relations among Supreme Court justices following a shocking allegation of an altercation in the Court's chambers. Mediation was accepted and successful in one of those cases. It may be that the convergence of the current political climate and confidence in mediation indicate an increasing role for mediation in the public arena.

Whereas the potential for seemingly unproductive episodes and impasses in political decision-making is inherent in representative democracy, it may be that mediation can be useful amid agreement on the need to overcome sustained gridlock for the public good. ◆

GROUP FACILITATION:

IAF
INTERNATIONAL ASSOCIATION OF FACILITATORS

A RESEARCH & APPLICATIONS JOURNAL

Contributions of Caucusing and Pre-Caucusing to Mediation

Gregorio Billikopf-Encina

Abstract:

Drawing on his work as a researcher and practicing mediator in interpersonal organizational conflict, the author argues that pre-caucusing—a separate meeting between the mediator and each of the stakeholders *before* they are ever brought together into a joint session—can not only overcome many of the negatives often associated with caucusing, but has the potential of becoming a pillar of conflict management. This is especially so when pre-caucusing is integrated into a transformative mediation framework. Pre-caucusing affords stakeholders the opportunity to vent and be heard at a critical time in the mediation process, when it can reduce defensiveness and increase creativity. Once in the joint session, stakeholders communicate with each other with less mediator interference.

Key Words:

caucusing, pre-caucusing, mediation, conflict resolution, alternative dispute resolution, conflict management skills, conflict management mechanics

Contributions of caucusing and pre-caucusing to mediation

Wherever choices exist, there is potential for disagreement. Such differences, when handled properly, can result in richer, more effective, creative solutions. But alas, it is difficult to consistently turn differences into opportunities. When disagreement is poorly dealt with, the outcome can be *contention*. Contention creates a sense of psychological distance between people, such as feelings of dislike, alienation, and disregard. Such feelings can get in the way of effective communication and resolution of even the most minute perceived differences (Billikopf 2000).

Deep-seated interpersonal conflict requires an enormous amount of skill to mediate, even when the best of present-day theory is put into practice by trained and skilled mediators. Yet others who may have little mediation training, such as facilitators, may at times find themselves in the role of mediator.

Despite years of experience as an admired and skillful facilitator, a colleague confessed that mediation required specialized skills. He described a recent intervention as a third-party neutral, one where he felt thrown into a lion's den. The stakeholders became involved in an ugly escalation right in front of him. As a mediator he felt impotent to help, and was even threatened by one irate stakeholder.

There are a number of subtle differences between what facilitators and mediators do. Although they both draw from a subset of common tools, there are important distinctions. Generally speaking, facilitators tend to help groups through the process of problem solving and creative decision making. Mediators often deal with stakeholders who may be more openly antagonistic toward each other.

Facilitators, in many cases, work with situations where people may not know the way, but are excited about finding a common direction. Mediators, in contrast, often work with those who

have lost faith in the other stakeholder, as well as any hope of resolving the challenges in a mutually positive or amicable fashion. Having made such broad generalizations, it is important to note that individual mediators and facilitators vary enormously both in philosophy and approach.

There are times when interpersonal conflict may force a facilitator to concentrate on individual or group antagonisms. At times like this, the facilitator may benefit from additional mediation skills.

The focus of this paper is on the contributions of *caucusing* as a mediation tool, and more specifically, the use of *pre-caucusing*. In caucusing, the third-party neutral meets separately with each stakeholder, in the absence of the other contending party. In pre-caucusing, these separate meetings take place *before* the mediator ever brings the stakeholders into a joint session (Billikopf 1994; Billikopf 2000).

While countless factors are involved in successful mediation, some are so compelling that they may be called *pillars* of mediation. Pre-caucusing may well be such a pillar.

With notable exceptions, caucusing has received a somewhat uneven and often shallow treatment in the literature. Little is said explicitly about pre-caucusing. Certain value assumptions about mediation further complicate some of the controversy surrounding the topic. One of the most important of these values involves mediator choice between a *transformative* (Bush & Folger 1994) and the more traditional *directive* mediation.

The *directive* approach tends to focus on finding an acceptable agreement—one that may involve *settling* or *compromising*—between the contending parties. It is sometimes called directive because of the large amount of power and responsibility placed on the mediator. Some mediators may come close to acting as arbitrators, imposing a solution on the participants. Of course, mediators do not normally start out thinking that they will impose a solution. As situations become more difficult and emotional, however, it is increasingly likely that directive tactics will be utilized (Bush and Folger 1994; Folger, Marshall, & Stutman 1997; Lewicki *et al.,* 1994).

Transformative mediation

• allows stakeholders to retain maximum control over the process

• creates an atmosphere where parties can begin to connect interpersonally (i.e., provide mutual recognition or support)

• helps parties become better negotiators and reduce dependence on neutrals

• seeks solutions that are based on a careful understanding of the problem, rather than rushing into agreements that may be short-lived.

A study on self-esteem found that people prefer conflict management situations in which they have added control over the results, even when such control may mean making greater concessions (Swann 1996). My own preference towards transformative mediation affects how I see and utilize caucusing.

We shall first review what is said about pre-caucusing in the literature. The positive and negative attributions often associated with caucusing, and particularly, the special contribution played by pre-caucusing, are mentioned next. Examples of pre-caucusing are drawn from my involvement as a researcher and mediation practitioner in organizational settings.

Pre-caucusing in the Literature

Little is said in the literature about either pre-caucusing or the timing of caucusing in general. For instance, Moore suggests, "Mediators should take care not to schedule caucuses prematurely, when parties are still capable of working productively in joint session, nor too late, after unproductive hostile exchanges or actions have hardened positions" (1996, p. 320).

Bush and Folger are more explicit about the benefits of early caucusing: "Exploring delicate relational issues and laying further groundwork for recognition is sometimes easier in caucus, especially in the early stages of the process. Parties often find it difficult at first to give recognition directly to the other party, because it is difficult to give recognition to another person when feeling vulnerable oneself" (1994, p. 153). Having said that, however, they warn that breaking into caucus too early may interrupt the "transformative momentum" or positive conversation flow between stakeholders that may involve positive acts of mutual recognition (Bush & Folger 1994, p. 271).

There is one veiled reference to pre-caucusing, mentioned almost as an aside by Folger, Marshall, and Stutman. In a sidebar case, a mediator was using computer technology as an aid to conflict resolution. The mediator is reported to have met with the stakeholders "separately prior to the session to help them clarify their needs and positions" (1997, p. 285).

Volkema comes close to suggesting a pre-caucus: "The first contact between the mediator and the parties provides the first opportunity to establish public images. If this contact is between the mediator and one other person, only two identities need to be negotiated, although groundwork for others can be laid at the same time" (1988, p. 8).

Winslade and Monk (2000) are clear proponents of the pre-caucus, especially in cases involving entrenched disputes, although they studiously avoid the word caucus, given its negative associations:

> One of the first steps we prefer to take in a mediation is to meet with each of the parties separately …. In our experience, it is in these separate meetings that a lot of the major work of the mediator is done. … the separate meetings are a venue for significant developments in the mediation as a whole, not an optional adjunct to the process, to be used only when things are getting sticky. In our approach, they are central to what gets achieved (2000, p. 137).

Despite Winslade and Monk's use of the pre-caucus, I found they failed to take advantage of all of the pre-caucus's transformative possibilities. In the joint session stakeholders tend to address the mediator rather than each other. In fairness to Winslade and Monk, this happens even in the approach used by Bush and Folger (1994).

POSITIVE CONTRIBUTIONS OF CAUCUSING

Positive attributes usually associated with caucusing include:

*   deciding whether to bring the parties together into a joint session (Moore 1987; Moore 1996)

*   giving the opportunity to stakeholders to vent (Blades 1984; Emery & Jackson 1989; Hobbs 1999; Hohlt 1996; Moore 1987; Moore 1996; Pruitt *et al.* 1989; Welton, Pruitt, & McGillicuddy 1988)

*   helping each party feel understood by the mediator (Emery & Jackson 1989; Hobbs 1999; Hohlt 1996; Moore 1987; Moore 1996; Pruitt *et al.* 1989; Volkema 1988; Welton, Pruitt, & McGillicuddy 1988)

*   exploring positions and needs (Blades 1984; Castrey & Castrey 1987; Emery & Jackson 1989; Hobbs 1999; Hohlt 1996; Moore 1987; Moore 1996; Pruitt *et al.* 1989; Volkema 1988; Welton, Pruitt, & McGillicuddy 1988)

*   reminding parties of the benefits of mediation (Moore 1987; Moore 1996; Volkema 1988)

*   coaching stakeholders on effective communication and negotiation techniques (Hobbs 1999; Moore 1987; Moore 1996; Volkema 1988)

*   appealing to stakeholders' higher principles (Blades, 1984; Hobbs 1999; Hohlt 1996; Moore 1987; Moore 1996; Pruitt *et al.* 1989; Volkema 1988; Welton, Pruitt, & McGillicuddy 1988; Winslade & Monk 2000).

Each of the next several sections (1) presents a key decision or outcome of mediation, then (2) underscores the contributions of caucusing followed by (3) the additional benefits of pre-caucusing.

Deciding to Bring Parties Together

The ideal is to bring the stakeholders together so they can make a joint decision and retain maximum control over the situation. An important outcome of effective mediation is to enable parties to handle future challenges without a mediator.

While the results of mediation can be markedly superior to those obtained through other third-party interventions (such as arbitration), this is not necessarily so with substandard mediation (Castrey & Castrey 1987). When things go wrong in mediation, stakeholders may take advantage of the sense of safety they feel in order to escalate the contention to even higher levels than before. It is possible that the mediator can do *more harm than good* by bringing the parties together.

Contributions of caucusing

Moore suggests that a mediator may use caucusing to deal with relationship problems, and that at times a neutral third party may want to "discourage or prevent the parties from returning to joint session ... when extremely strong emotions [might] be a major stumbling block to further negotiations" (1987, p. 88).

Further contributions of pre-caucusing

A central aim of the pre-caucus is for the mediator to assess the potential benefits and harm of bringing stakeholders together, *before* any damage is done. When contention is allowed to come into the mediation session, the opportunity for stakeholders to start with a clean slate is compromised. Emotional escalation, as Moore (1987) suggests, may also have a negative effect on reaching agreement.

In one of my early efforts as a mediator, a manager not only refused to look at his assistant in the joint session, but turned his chair so as to present his back to her. After this experience I developed a litmus test to better help me gauge the likelihood that a joint session would be successful: asking a stakeholder for what he or she values in the other (Billikopf 2000). This question is so telling because people involved in deep-seated conflict may have trouble finding anything positive to say about another (Bush & Folger 1994). This is not a question to ask at the outset, as stakeholders may be in too much pain to see very clearly. Nor should the mediator take the first negative expression as final. (For additional tests see Lewicki, *et al.* 1994, p. 360-361.)

In one difficult case, a top manager could not make a single positive remark about a middle manager who worked for him,

despite the positive things that had been said about him. I shared with the top manager my experience that there was little likelihood of mediation success where an individual could find nothing positive to say about another, and suggested a short break. When we resumed our conversation, the recalcitrant manager was waiting for me with a list of sincere, positive feelings about the other stakeholder.

Opportunity to Vent

> Two couples sat on either side of the table, glaring hostilely at each other. At the head of the table, a schoolteacher in her thirties was explaining the service. "First you, Mr. and Mrs. A, will have a chance to tell your side of the story and Mr. and Mrs. Z will listen quietly. Then you, Mr. and Mrs. Z, will have the same opportunity. After that we will discuss the situation and try to find a way to resolve it ... . While each side was telling its story, there were outbursts from the other of "that's not true" or "wait a minute," which the mediator strove to contain. (Pruitt *et al.* 1989, p. 202)

Mediators often struggle unsuccessfully to maintain control over conflict escalation. Early joint session phases where stakeholders share their stories, come up with ground rules, or begin to interact, frequently lead to unconstructive exchanges. "After each parent has voiced concerns, the two parents are encouraged to discuss the issues freely. In the majority of cases, an argument ensues," say Emery & Jackson, who discuss child custody disputes. "The fight is almost always unproductive ..." (1989, p. 6).

Kenneth Kressel explains that it is a "common theme in the mediation cannon" (p. 25) to let each party tell their side of the story in front of the other. He then shares the destructive effect of this approach: "Mrs. Smith would accept my invitation [to tell her side of the story] with relish, explaining that they were here because Mr. Smith was a worthless lout who cared nothing for his children or common decency and had been vilifying and humiliating her for years. For all she knew, he might also be an alcoholic and child abuser. His cross-dressing was a matter of record. She was in mediation by order of the court and was certainly willing to do her best to encourage Mr. Smith to 'finally be a father' but was, shall we say, skeptical. Whatever the tonic benefits of this outburst for Mrs. Smith, for Mr. Smith and myself the results were clearly unhappy: he would be provoked into an apoplectic rebuttal and I into a dismal contemplation of other lines of work. Yes, I exaggerate. But only a little" (1994, p. 26).

Some mediators feel that such loss of control is unavoidable, part of the process, or even necessary (Emery & Jackson 1989, Rothman 1997). I contend, however, that there is a better way; that stakeholders have already experienced what does not work, and remember it well. It is hardly necessary for them to re-experience it now in front of the mediator. Most third-party neutrals would probably welcome an approach where such dysfunctional escalations were either greatly reduced or completely eliminated.

Some have suggested strategies for reducing such futile outbursts, including telling one party to remain silent or focus on listening (Hobbs 1999) while the other speaks. To make the point, the listening party may be given a notepad and asked to take notes (Emery & Jackson 1989). It has also been suggested that joint sessions be held in a public place to help stakeholders tone down their emotions (Folger, Marshall, & Stutman 1997). While the note-taking suggestion has some merits, in this context such artifacts may delay contentious outbursts rather than prevent them.

Contributions of caucusing

Stakeholders may have some very poignant and deeply antagonistic feelings towards each other. When these can be vented in front of the mediator, the stakeholder often has less of a need to vent in a destructive manner in front of the opposing stakeholder. Defensiveness is reduced and creativity increased as the mediator protects stakeholders from further mutual abuse.

There is little disagreement on this point: while involved in caucusing stakeholders are less hostile than in joint sessions (Welton, Pruitt, & McGillicuddy 1988). When conflict escalates into contentiousness, as in these episodes, the mediator not only permits stakeholders to lose face, but just as importantly, she or he loses both control (Butler 1994) and face (Volkema 1988) in front of the stakeholders.

Further contributions of pre-caucusing

When dealing with acquaintances or strangers, individuals often go out of their way to make an effort to project their best possible behavior. This is especially true in what could be called a "courting period." This honeymoon period may last years, when stakeholders view their relationship as fair and equitable. When the rules of proper interpersonal exchange are violated (Brown, 1986) and someone feels taken advantage of, the situation can change quickly.

Similarly, in a stakeholder's relationship with a mediator—assuming the mediator is a stranger and/or has the respect of the stakeholders—individuals often try extra hard to be on their best behavior (Folger, Marshall, & Stutman 1997), lest the mediator think that they are culpable. Stakeholders are more likely to want to continue to make a good impression on the mediator after they have established themselves as reasonable people in the pre-caucus. Volkema suggests that "it is not unlikely that the parties will have established one image with each other and another image with the mediator" (1988, p. 11).

People also attempt to be consistent: "Consistency gives actors a desirable degree of predictability and trustworthiness, and it generates liking and respect" (Schlenker 1980, p. 232). Stakeholders are likely to feel a greater need to be seen as consistently reasonable by a mediator who has had sufficient time to meet with them individually. Effective listening is a very powerful tool, and people tend to respect those mediators who can listen with care and empathy.

Once the parties have exchanged insults in front of a third-party neutral in traditional mediation, on the other hand, much of the damage has been done. Stakeholders feel less motivated to show their best after exposing their worst behavior.

It is not that stakeholders pretend to be someone they are not. Because stakeholders meeting with the mediator in the pre-caucus know they will be meeting with the other party in a joint session, it is my experience that they are likely to share their own shortcomings, rather than wait for the other party to bring these out. It is this new facework (allowing another to save face) between stakeholders that the mediator wants to encourage in order to give parties an opportunity for a fresh start that is not based on blame.

Helping Each Party Feel Understood by the Mediator

It is difficult to expect stakeholders who have been involved in deep-seated conflict to put aside their own needs and listen to and focus on the needs of the other party (Bush & Folger, 1994). The natural tendency is for stakeholders to want to express their own perspectives first. The more deep-seated and emotional the conflict, the greater this tendency.

At times, tension in deep-seated interpersonal conflict situations can reach almost unbearable levels. In mediating such conflicts within organizations, it is common for stakeholders to strongly contemplate withdrawal from the enterprise. Psychological separation from the other stakeholder and possibly from the organization has already taken place. For instance, in child custody mediation, parties have already separated physically and psychologically from each other, yet need to work together for the benefit of the children involved.

Contributions of caucusing

Because stakeholders have the opportunity to meet separately with the mediator, each gets the opportunity to explain his or her perspective first, before having to attend to the other participant. When the stakeholder feels understood, an enormous emotional burden is lifted, thus making him or her more receptive to listen to others (Covey 1989). It is true that stakeholders have a special need to be understood by the other party in the contention, but being understood by the mediator contributes much. Often, it is a necessary step in terms of a stakeholder gaining enough confidence to proceed further.

Some individuals tend to be more silent than others. Caucusing increases the chances that an individual will talk (Hohlt 1996) and express his or her feelings. It is hardly possible for the mediator to help individuals who refuse to speak about "where it hurts." Mediators have the opportunity to show empathy in a caucus situation without arousing jealousies in other party.

Further contributions of pre-caucusing

It is at the start of mediation that stakeholders are perhaps most apprehensive as to what mediation may bring. Parties often come to the table with every defensive mechanism armed and ready to deploy (such as sulking silence, angry outbursts, combative body language). They may have trouble looking at the mediator, let alone the other party.

When a pre-caucus is used and the other party is not present, this stakeholder frustration and despair is re-directed in more positive ways. To have an empathic ear to listen to a stakeholder in such a non-judgmental way is powerful medicine indeed. I have seen people who were supposed to be "silent types" open up and talk freely. Men and women have wept openly as they released tension. Such emotional releases are not available to stakeholders in more traditional mediation.

The Exploration of Needs and Benefits of Mediation

The mediator attempts to understand individual items under dispute, as well as the general perspectives of stakeholders, and helps stakeholders keep alive the benefits of mediation (in contrast to other alternatives such as arbitration).

Contributions of caucusing

An important benefit of caucusing is being able to explore beyond positional bargaining, into stakeholder interests and needs (Fisher, Ury, & Patton 1991). Stakeholders can also be reminded that mediation confers tangible benefits over interventions where they have less control. This is more likely to happen when individuals feel less vulnerable and defensive, and are more willing to think aloud without feeling forced into making concessions. A mediator can increase her or his understanding of the situation through such exploration, but more important yet, the self-awareness of each stakeholder increases. For instance, it may become clear that a stakeholder desires an apology, rather than some other remedy.

Further contributions of pre-caucusing

When disputants enter the joint session with the benefit of a pre-caucus, the mediator can often take a less visible role. Each stakeholder comes to the joint session possessing enhanced clarity about the issues and self-confidence.

In one situation, after I listened to stakeholders during a pre-caucus, they were able to go on and solve the problem on their own. Bad feelings had developed between them concerning how each introduced the other to visitors and the media. Not only did they solve this problem on their own; they also dealt with related underlying issues, and even went on to discuss opportunities for future career growth and cooperation (Billikopf 2000).

As a neutral I sometimes do little more than introduce topics brought up during the pre-caucus. Allowing the stakeholders to solve an easier problem early on may give them the needed boost to deal with more challenging issues later (Blades 1984; Emery & Jackson 1989). Furthermore, a mediator who understands the issues involved can make sure that significant issues are not ignored. Despite previous antagonisms, communication between stakeholders during joint sessions is sometimes so fast paced that I have to scramble to understand and note their agreements. At times like these I feel like an unneeded observer. Setting up a situation where stakeholders address each other with little mediator interference takes transformative mediation to the next level. Although not all cases achieve this ultimate success, mediators can count on better communication flow and reduced contentiousness between stakeholders.

Educate Stakeholders on Effective Negotiation Skills

One measure of mediation success is when it equips stakeholders to handle future challenges on their own. While this may not necessarily happen after a single experience with mediation, the stakeholders can take with them increased self-awareness and conflict management skills.

Stakeholders may be shown how they can present a perspective using neutral or non-provocative language (Hobbs 1999) and without causing the other to lose face. An important part of conflict management is helping stakeholders recognize the need for the other party to build and save face (Ting-Toomey 1999; Volkema 1988; Blades 1984; Moore 96). In the absence of these skills, people are likely to revert to a more dysfunctional and emotional approach to communication. Participants may also develop a better understanding of the nature of conflict — learning how to divide big issues into smaller ones and what constitutes a proper apology, for instance. Both stakeholders gain negotiation power as they improve their ability to communicate in effective ways.

Contributions of caucusing

Mediators have the opportunity to privately discuss participant behaviors that are working, as well as those that are not. This avoids the appearance of favoritism associated with public compliments, as well as the loss of face connected with open criticism.

Further contributions of pre-caucusing

It is hard to expect the stakeholders to have a positive mutual conversation when they lack even the most rudimentary notion of how their communication strategies affect the other stakeholders. Those who grasp new insights into the negotiation process early on are more likely to enter the joint session feeling confident and prepared, with some control over the results.

Among the potential positive outcomes of transformative mediation is giving parties the opportunity to apologize and to accept an apology (Bush & Folger 1994). One stakeholder had a history of vitriolic temper outbreaks when I first met with him. His anger often manifested in shouting and profanity. During the pre-caucus, it became increasingly clear that this stakeholder felt no regret about his temper tantrums. He was quick to both minimize the extent of his anger, and to justify his bullying behavior. Had he defended such behavior in a joint session, his credibility would have been greatly damaged. Through a series of role-plays and conversations during the pre-caucus, he came to understand the importance of offering an apology for his profanity and anger. Furthermore, he suggested that the topic be brought up early on in the joint session so he could have a chance to apologize. During the first role-play his words had sounded shallow at best. The actual apology offered during the joint session was moving and sincere.

Regular caucusing has one advantage over pre-caucusing here. While the mediator can observe and coach a stakeholder during a pre-caucus, some dysfunctional communication approaches only manifest themselves during the joint session. This is not a fatal flaw of pre-caucusing, because a regular caucus can be utilized later to deal with such issues.

Much of what has been said here also applies to the idea of appealing to a stakeholder's higher principles. Many transformative opportunities that could otherwise be lost present themselves during the pre-caucus. For instance, an owner-operator said something touchingly positive about one of his managers during the pre-caucus. I suggested that it would be magnificent if he could share that thought with the other stakeholder during the joint session. The owner explained that he would never do so. I challenged him to reconsider, but left the ultimate decision up to him. The individual chose to share the affirming comment during the joint session, taking ownership for that decision, thus making it his own.

NEGATIVE CONNOTATIONS OF CAUCUSING

A number of challenges are associated with caucusing, including:

- lack of stakeholder truthfulness (Pruitt *et al.* 1989; Volkema 1988; Welton, Pruitt, & McGillicuddy 1988)

- mediator bias (Blades 1984; Engram and Markowitz 1985; Moore 1987; Moore 1996; Pruitt *et al.* 1989; Volkema 1988; Welton, Pruitt, & McGillicuddy 1988)

- mediator control or abuse of power (Blades 1984; Folger, Marshall, & Stutman 1997; Keltner 1965; Moore 1987; Moore 1996; Pruitt et al. 1989; Volkema 1988)

- reduced likelihood that disputants will know how to handle future challenges (Pruitt et al. 1989)

- mediator violation of confidentiality (Blades 1984; Moore 1987; Moore 1996)

- interruption of positive movement (Moore 1996; Welton 1988)

- free time for the other stakeholder to use in an effort to build his or her own case (Welton 1988).

Attacks on Directive Mediation

As we shall see, most criticisms associated with caucusing are really attacks on directive mediation, rather than on caucusing itself. Where caucusing is instead used to increase stakeholder control through transformative mediation, most of these objections melt away.

As positive as mediator empathy toward a stakeholder may be, some fear this may lead to stakeholder untruthfulness. They reason that the absence of the other party during the caucus leaves the stakeholder free to exaggerate. Others argue that caucusing may lead to deals between the neutral party and one of the stakeholders. "Disputants often fear that clandestine deals or coalitions [may take place] between the other party and the mediator" (Moore 1996, p. 200).

Yet others suggest that caucusing simply gives the mediator too much control, lends itself to abuse of mediator power, and does little to equip stakeholders for future conflict in life. Instead, they argue, stakeholders may become more dependent on mediation. "Caucuses ... are explicit attempts to narrow issues, to push for compromise, and to synthesize arguments and positions." (Folger, Marshall, and Stutman 1997, p. 262). We even read, "caucuses provide mediators with the greatest opportunity to manipulate parties into agreement" (Moore 1996, p. 325). Volkema warns that mediators with a vested interest

may promote one outcome over another (1988). The assumption, in all these cases, is that agreement is reached during caucusing.

There is nothing inherent in caucusing itself, however, that leads to these difficulties. Quite the contrary, Engram and Markowitz suggest that "... the judicious use of caucusing in ... mediation can even enhance the perception of neutrality and will result in increased trust in the process of mediation" (1985, p. 25). Likewise, where transformative mediation is used, caucusing may be seen as a tool to help stakeholders become better negotiators (Bush and Folger 1994).

In transformative mediation where it is *the stakeholders who solve their own disputes*, there is little to be gained by attempts to influence the mediator. Stakeholders need not be concerned that the mediator will make a secret agreement with the other stakeholders. Caucusing is used to teach negotiation skills to stakeholders, rather than to circumvent stakeholder empowerment.

Violation of Confidentiality

Another negative associated with caucusing is the potential for sharing confidential information obtained from one stakeholder, either purposely or through a slip. Certainly, mediators need to be careful not to divulge confidential information. Yet it should be clear that the purpose of caucusing is to help stakeholders better understand their own needs and prepare to communicate these to the other party in the joint session—not to talk about issues stakeholders want to keep secret from the other participant. True, some subjects are originally brought up in a somewhat raw manner. These are translated into more effective messages that tend to reduce defensiveness. For instance, if a stakeholder feels the other is inconsiderate or selfish, the mediator helps the stakeholder better understand critical incidents that may have led to this evaluation. During the joint session, the incidents and behaviors are discussed without the labels.

As a mediator, I note all the issues that are important to stakeholders during the pre-caucus, and give them a chance to expose these during the joint session: "A, Could you share with B the story you told me about X." Opportunities are balanced for both stakeholders to bring up issues that are then jointly discussed.

Sometimes ethical issues require disclosure, such as when a spouse is hiding an asset from the other during a divorce settlement. In those situations, Blades (1984) suggests that the mediator make it clear to the pertinent stakeholder that the neutral's continued involvement in the mediation depends on the stakeholder disclosing this information to the other party. Standards have been suggested for issues with and limits to confidentiality (Milne 1985; Moore 1987). Caucusing does not

cause an inherently unethical situation to develop, however. It simply affords the mediator an opportunity to help correct an unfair situation. "Much of the controversy surrounding the issue of caucusing ... stems from differences in training or orientation rather than from a real debate about ethics" (Engram and Markowitz 1985, pp. 24-25).

Interruption of Positive Movement

Caucusing may be called at any time, by stakeholders or by the mediator. Stakeholders may even wish to caucus within their own team, without the mediator. Alternatively, the mediator may need time alone and call a "mediator caucus" (Castrey and Castrey 1987, p. 15). Any type of caucusing may interrupt the positive flow of the conversation. The great advantage of pre-caucusing is that it does not interrupt the positive flow of communication that may be established during the joint session. Furthermore, pre-caucusing probably reduces interruptions after the joint meeting has begun.

Free Time to Solidify Stance

The concern that caucusing permits one party time to further solidify her or his own stance while the other is engaged in caucusing, is simply a non-issue. In transformative mediation one of the roles of the mediator is to help stakeholders consider potential pitfalls. Mediators help stakeholders truly understand the problem and thus avoid quick unworkable solutions.

CONCLUSIONS

Contention creates a sense of psychological distance between people, turning even minute differences into ones that seem insurmountable. A tool of particular value is the caucus, where the mediator meets separately with stakeholders. The literature has shed light into both the positive and negative contributions of caucusing. Positive aspects of caucusing include giving parties an opportunity to tell their story and be heard, explore needs, and vent privately. Mediators may also take advantage of caucusing to coach parties and help them understand the tools that will help them become better negotiators in the future.

Interestingly, most of the criticisms associated with caucusing derive from a directive mediation approach. When caucusing is used within a transformative framework, most of these potential negatives disappear. In transformative mediation the stakeholders remain the primary actors. Not only do the contending parties retain control over the outcome, they are also equipped with many of the tools they will need to solve future problems: "A skillful transformative mediator can use caucuses in a manner that not only avoids the problem-solving pitfalls [found in the directive approach] but actually builds transformative momentum over the course of a session" (Bush and Folger, 1994, p. 270).

Although in the literature we find some allusions to the benefits of the pr-caucus, very little is said explicitly about it. When pre-caucusing is used with a transformative approach to mediation, the benefits of caucusing are multiplied, and the potential negatives are further reduced.

The main reason why pre-caucusing is effective is that the mediator affords each stakeholder the opportunity to be heard when he or she needs it the most. A conflict situation that calls for mediation, almost by definition, is a difficult one. Stakeholders are most often focused internally and have little capacity to listen to someone else at the beginning of mediation. This internal focus tends to extinguish creativity by increasing negative emotion and defensiveness. A stakeholder who feels heard in the pre-caucus is better able to listen to the other stakeholder and to connect in a more positive way. The groundwork laid out during the pre-caucus allows stakeholders to address each other with little mediator interference.

Mediation has the potential to do much good. Poorly carried out mediation, where contenders feel they can exchange insults in a psychologically safer environment, can do more harm than other forms of neutral-party interventions. The pre-caucus affords mediators the opportunity to make difficult decisions as to whether to bring contenders into a joint session.

Sometimes the most productive approaches are the simplest, and this is certainly true with the pre-caucus. Caucusing as a mediation tool has been partially misunderstood and certainly has not been used to its potential.

References

Billikopf, G.E. (1994). *Labor Management in Agriculture: Cultivating Personnel Productivity.* Modesto: University of California Agricultural Extension.

Billikopf, G.E. (2000). Ch. 13. Conflict Management Skills [Online]. *Labor Management in Agriculture: Cultivating Personnel Productivity.* Modesto: University of California Agricultural Extension. http://www.cnr.berkeley.edu/ucce50/ag-labor/7labor/13.htm

Blades, J. (March 1984). Mediation: an old art revitalized. *Mediation Quarterly* (3), 59, 78-79.

Brown, R. (1986). *Social Psychology: The Second Edition.* New York: Free Press.

Bush, R. and Folger, J. (1994). *The Promise of Mediation: Responding to Conflict Through Empowerment and Recognition.* San Francisco: Jossey-Bass Publishers.

Castrey, R.T. and Castrey, B.P. (Summer 1987). Timing: a mediator's best friend. *Mediation Quarterly*(16), pp. 15-19.

Covey, S. (1989). *Seven Habits of Highly Effective People.* New York: Simon & Schuster.

Emery, R.E. and Jackson, J.A. (Summer 1989). The Charlottesville mediation project: mediated and litigated child custody disputes. *Mediation Quarterly*(24), pp. 3, 6-7.

Engram, P. and Markowitz, J.R. (June 1985). Ethical issues in mediation: divorce and labor compared. *Mediation Quarterly* (8), 19, 24-25.

Fisher, R., Ury, W., and Patton, B. (1991). *Getting to Yes: Negotiating Agreement Without Giving In* (2nd ed.). New York: Penguin Books.

Folger, J., Marshall, P., and Stutman, R. (1997). *Working Through Conflict: Strategies for Relationships, Groups, and Organizations* (3rd ed.). New York: Longman.

Hobbs, K.S. (November 1999). Attention attorneys! How to achieve the best results in mediation. *Dispute Resolution Journal*(4), pp. 43-47.

Hohlt, E.M. Effective Use of Caucus, in Trachte-Huber, E. and Huber, S. (1996). *Alternative Dispute Resolution: Strategies for Law and Business.* Cincinnati, Ohio: Anderson Publishing Co., pp. 335-337.

Keltner, J. (1965). Communications and the labor-management mediation process: some aspects and hypotheses. *Journal of Communication*, p. 74. (In Moore, *The Mediation Process*, 1996)

Kressel, K. (1994). Frances Butler: Questions that lead to answers in child custody mediation. In Kolb, D.M. (Ed.), *When Talk Works: Profiles of Mediators.* San Francisco: Jossey-Bass Publishers, pp. 17-58.

Lewicki, R., Litterer, J., Minton, J., and Saunders, D. (1994). *Negotiation* (2nd ed.). Bur Ridge, Illinois: Irwin.

Milne, A. (June 1985). Model standards of practice for family and divorce mediation. *Mediation Quarterly* (8), pp. 73, 77.

Moore, C.W. (1996). *The Mediation Process: Practical Strategies for Resolving Conflict* (2nd ed.). San Francisco: Jossey-Bass Publishers.

Moore, C.W. (Summer 1987). The caucus: private meetings that promote settlement. *Mediation Quarterly*(16), pp. 87-101.

Pruitt, D.G., Fry, W.R., Castrianno, L., Zubek, J, Welton, G.L., McGillicuddy, N.B. and Ippolito, C. (1989). In M.F. Afzalur (Ed.), Ch. 15. *The process of mediation: caucusing, control, and problem solving*. NY: Praeger Publishers, pp. 201-208.

Rothman, J. (1997). *Resolving identity-based conflict in nations, organizations, and communities* (2nd ed.). San Francisco: Jossey-Bass Publishers.

Schlenker, B. (1980). *Impression Management: The Self-Concept, Social Identity, and Interpersonal Relations.* Monterey, California: Brooks/Cole Publishing Co.

Swann, W. (1996). *Self-Traps: The Elusive Quest for Higher Self-Esteem.* New York: Freeman.

Ting-Toomey, S. (1999). *Communicating across cultures.* New York: The Guilford Press.

Volkema, R.J. (Winter 1988). The mediator as face manager. *Mediation Quarterly* (22), pp. 5-14.

Welton, G.L.; Pruitt, D.G. and McGillicuddy, N.B. (March 1988). The Role of Caucusing in Community Mediation. *Journal of Conflict Resolution,* 32(1), pp. 181-201.

Winslade, J., and Monk, G. (2000). *Narrative Mediation: A New Approach to Conflict Resolution.* San Francisco: Jossey-Bass Publishers.

Gregorio Billikopf-Encina is the author of *Labor Management in Agriculture: Cultivating Personnel Productivity.* Billikopf received his Bachelor of Science in Plant Science from UC Davis, and his Masters in Human Resource Management from California State University, Stanislaus. Since 1981, he has been employed with the University of California Agricultural Extension as a Labor Management Farm Advisor. Gregorio has been a frequent national and international speaker in his field, including presentations in Russia, Canada, Uganda, México, and his native Chile. Research and teaching emphases have been in the areas of employee selection and testing, incentive pay and pay issues, conflict resolution, employee discipline, performance appraisal, worker motivation, supervision, negotiation skills and interpersonal relations. Contact at University of California / 3800 Cornucopia Way # A / Modesto, CA 95358-9492; (209) 525-6800, Fax (209) 525-6840, gebillikopf@ucdavis.edu

Raising Awareness of Potential Biases and Microaggressions: Applications to Mediation

Alan E. Gross

Abstract

"Neutrality" or "impartiality" are well accepted and practiced values that are important cornerstones of mediation; nonetheless mediators usually bring biases and preferences with them into the room, as do the parties. Mediators are responsible for becoming aware of these biases, especially when they are unconscious, so that they can conduct themselves in an impartial manner. Mediators sometimes can encourage parties to bring their biases into the open where they can be dealt with more effectively. Especially difficult kinds of insulting behavior are "microaggressions" defined as minor slights that are usually well-intended by parties and mediators. This article is intended to raise awareness of biases and microaggressions, and to suggest means for preventing them and ways of responding to situations in which they commonly occur.

Keywords

Microaggressions are small verbalizations or other behaviors which, intentionally or not, are perceived as insulting, demeaning or biased (e.g., racist, sexist, homophobic) by the victim or target person. For example: "you are a credit to your (race, gender, etc.)" may communicate that the target group is not as intelligent as (whites, men, etc.). Because majority culture people, including mediators, are often unaware that they are sending negative messages, sensitizing them to the effects of such behavior on targets can considerably reduce such unwanted behavior.

The term "microaggression" was introduced by Charles Pierce in the 1970s (Pierce et. al., 1978). Pierce's concept was largely based on and confined to subtle and not-so- subtle racism in broadcast television in the 1970s and earlier. However, in the early 21st century, Prof. Derald Sue and his colleagues have identified microaggressions in many settings, and have broadened the concept to include sexism, homophobia and much more. Sue's book "Microaggressions in Everyday Life" (Sue, 2010) is the definitive and comprehensive work in the field. Sue and his colleagues describe m a n y varieties of microaggressions as presented in a taxonomy in their seminal American Psychologist article (Sue et. al. 2007). Among the examples quoted in that article: "You speak good English" (assumption that an Asian

or Latino American is foreign born); "You are a credit to your race" (it is unusual from someone of your race to be intelligent); "When I look at you I don't see color" (you should assimilate to the dominant culture; denying a person's racial/ethnic experiences); "I'm not racist. I have several black friends" (I am immune to racism because I have friends of color); "Everyone can succeed if they work hard" (People of color are lazy/incompetent and need to work harder).

Subcategories of Microaggressions such as Microinsults and Microinvalidations are usually related to race, ethnicity or gender but they can also occur as an (unintentional) expression of various "isms" including Ableism, Adultism, Ageism, Classism, Ethnocentrism, Heterosexism, Homophobia, Racism, Religious bigotry, Sexism, and Sizism. Many other concepts are related to Microaggression and biased treatment of others including bias, bigotry, culture, prejudice, scape-goating, and positive and negative stereotyping.

It should be noted that although Sue and many others introduced the microaggression concept in order to increase awareness among both potential perpetrators and victims so that harm to victims could be reduced or eliminated, a recent backlash has emerged from those who equate attempts to reduce microaggressions with "political correctness," or suppression of free speech as championed by the first amendment to the U.S. Constitution. When some American universities specified what kind of microaggressions were not acceptable in the classroom and elsewhere, those directives evoked especially intense negative reactions. See for example Prof. Volokh who accuses the University of California administration of suppressing viewpoints and even attacking academic freedom: https://www.washingtonpost.com/news/volokh-conspiracy/wp/2015/06/16/uc-teaching-faculty-members-not-to-criticize-race-based- affirmative-action-call-america-melting-pot-and-more

However, in the practice of mediation, and elsewhere in the social world, I fail to see a conflict between freedom of expression and the desire to prevent harm to victims of microaggressions. Considerate citizens already observe, often voluntarily, limits on what they express from not shouting "fire" when panic might result, to subtler restrictions such as naming ethnic, racial and other groups by the labels that they prefer. Mediators would likely come down on the side of protecting parties from bias ranging from intentional insults to subtle or unintended microaggressions. As Sue and others suggest, the first step in reducing the negative impact of such comments is to make them visible so that they can be unpacked, clarified, and discussed openly.

Defining and recognizing microaggressions is difficult but it is only a first step (Sue, 2010: 40). The mediator must then decide whether to intervene and how. Two basic kinds of microaggressions can occur at the mediation table: 1) one of the parties may unintentionally microaggress against another party 2) mediators may themselves become a perpetrator or, very rarely, a target of a microaggression. In the second case the perpetrating mediator who becomes aware that he/she has targeted one or several parties, especially if they react negatively, may effectively apologize for the unintended insult/slight. The first case is more problematic because a mediator intervention to protect the party who is the target of a microaggression, or to criticize or "educate" the perpetrator could compromise impartiality or provoke defensiveness. In such cases, an intervention in caucus might be preferable. In general, targets of

microaggressions are often conflicted about whether and how to respond. Pushing back in anger could exacerbate or lead to accusations of oversensitivity, but not responding may also be psychologically difficult. By recognizing and commenting on an microaggression, the mediator can possibly take the burden of responding from the target/victim. Parties who intentionally or even inadvertently express overt racism might expect a negative reaction; however, microaggressors are often surprised by negative reactions because their intention is often to compliment or relate to the target.

My focus here is to borrow ideas from Sue and others who write about microaggression and to apply them to what can happens at the mediation table. Although microaggressions do not frequently occur during mediation, mediation sessions are not immune from such subtle biases. A colleague at a New York City community mediation center recently described these two incidents, one non-verbal and one a verbal "compliment": 1) In a waiting room she observed a mediator warmly greet a white male party with a friendly hand shake, but only nodded a greeting at a black female party. 2) In an emotional session involving an African American mother and her teen son, a white middle-aged female mediator addressed the son with "You're a tall good looking black youth. Why not try some athletics or a part-time job?" These two well- intended recommendations had almost nothing to do with the content of the session (fighting at school). In the latter case, the co-mediator was aware of the microaggression, and called a recess to counsel her partner. One means of sensitizing mediators to such problems is to present them with hypothetical scenarios during training.

Here is one such handout that I have used at trainings.

1) You and your partner are mediating a dispute where one of the parties is black. You are white and your partner is black. At one point during the mediation, the black party looks at you and says that you cannot understand what he/she is dealing with, and then turns to your black partner and suggests that your partner must be able to understand his/her problem. *How would you handle this situation?*

2) You are co-mediating with partner of opposite gender. Parties are two men from Latin America who direct all their comments and eye contact to the male co-mediator. *How would you handle this situation?*

3) You are mediating a dispute between neighbors at a community mediation center. During a caucus, one party says that they don't get along because he is Puerto Rican and the other party is Dominican and hates Puerto Ricans. It appears to you that the complaints about noise and other problems in the relationship are very closely related to the Dominican party's dislike of the Puerto Rican party, and that the hatred arises out of their different background. *How would you handle this situation?*

4) You are mediating a dispute between neighbors at a community center. Based on comments made by Party A, it becomes clear that one of the main causes of the many problems that the two parties have is that Party A doesn't like Party B because Party B is gay. Party A cannot comprehend how or why anyone would lead such a "disgusting life". *How would you handle this situation?*

5) You and your partner are mediating a case involving a Muslim couple on one side. The wife never answers or makes eye contact. You secede to ask to speak to the wife alone. Husband visibly upset.

How would you handle this situation?

6) You are one of two middle age white women mediating a PINS case involving a black 7th grader and his parents. While the two parents appear comfortable speaking with you, the boy refuses to speak. You employ a variety of techniques in an attempt to engage him, yet nothing works. You sense that this young boy may not feel comfortable speaking openly with two middle aged white women, and this discomfort may be a significant factor in his silence. *How would you handle this situation?*

7) You and your co-mediator, who you are working with for the first time, are mediating at a community center. Your co-mediator is a white middle aged man and you are Asian.

At one point when the two parties are out of the room, he comments how much easier this mediation would be if the parties, one of whom is black and the other Latino, had been raised as Asians raise their kids. He further explains that Asian kids are obedient, hard-working and rarely cause trouble. You feel that he is doing a reasonably good job mediating, but are a bit put off by his comments. *How would you handle this situation?*

The following general recommendations may reduce or eliminate some of the problems noted here and in the training handout:

1) It is an important mediator responsibility and skill to acknowledge their own biases, but to keep them in check by displaying neutral behavior during a session.

2) If a bias affecting mediation becomes obvious, be transparent, own it, and recuse yourself if necessary.

3) There are occasions where parties need you to help focus their conflict on particular aspects of a dispute that may deal with bias or prejudice – so if there is an elephant or even a pesky mouse in the room, acknowledge it and ask the parties if it is an issue. Obvious isms should not be ignored or disregarded if they are a possible factor in mediation.

4) Sometimes the parties will want or need you to address an underlying aspect of the conflict but they are inhibited from stating it out loud. Be brave and say it out loud – allowing room to be wrong. Once an issue is said out loud, it may be useful to discuss it before moving on.

5) We tend to classify people that we do not know in categories such as gender, race, and age and assume that various characteristics cluster together. Only some of these assumptions and biases are true. Since assumptions are not always accurate, a better way to acquire knowledge about a person is to check out assumptions by asking questions, preferably open-ended.

In conclusion, microaggressions can easily find their way into the mediation room; however, mediator awareness can often prevent such slights.

References

Pierce, C. Carew, J., Pierce-Gonzalez, D., & Willis, D. (1978). An experiment in racism: TV commercials. In C. Pierce (Ed.), *Television and Education*. 62-68. Beverly Hills, California: Sage.

Sue, D.W., Capodilupo, C.M, Torina, G. C.Bucceri, J.M, Holder, A.M.B, Nadal, K.L.Esquilin, M. (2007) Racial microaggressions in everyday life: implications for clinical practice. *American Psychologist*, 62:4, 271-286.

Sue, D.W. (2010) *Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation*. Hoboken, New Jersey: John Wiley and Sons.

Alan E. Gross, Ph.D. has mediated, arbitrated, facilitated, and trained for over 25 years at many venues in the Northeast US and especially New York City where he has served as the Special Projects Coordinator and Board Director for the New York Peace Institute (formerly Safe Horizon) Mediation Program, and has previously acted as Senior Director, Training Coordinator and 9/11 Family Mediation Coordinator for Safe Horizon. That work with 9/11 victims was recognized with a US Department of Justice Volunteer for Victims Award.

He has also been appointed as ombudsman for the American Psychological Association and the New York Mayor's Action Center, as an arbitrator for AAA, FINRA, and attorney-client fee disputes and as mediator for the US Postal Service and the US Army.

Gross holds MBA and Ph.D. degrees from Stanford University and was formerly Psychology Professor and Department Chair at the University of Maryland. He is a Fellow of the Association for Psychological Science, and the author of a textbook and more than 50 chapters, articles and papers related to conflict resolution and social psychology. He is the recipient of the 2011 ADR Achievement Award from the Association for Conflict Resolution, Greater NYC Chapter, and teaches graduate courses in mediation skills at the NYU Center for Global Studies.

As a Founding Member of Mediators Beyond Borders, he has developed and delivered trainings and other assistance for Liberian refugees, Ghanaian attorneys and land settlement officers, UN special staff in Sierra Leone and a diverse country-wide Peace Network in Iraq.

The Role of Caucusing in Community Mediation
Author(s): Gary L. Welton, Dean G. Pruitt and Neil B. McGillicuddy
Source: *The Journal of Conflict Resolution,* Vol. 32, No. 1 (Mar., 1988), pp. 181-202
Published by: Sage Publications, Inc.
Stable URL: https://www.jstor.org/stable/174094
Accessed: 04-03-2020 16:39 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/174094?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

The Role of Caucusing
in Community Mediation

GARY L. WELTON
DEAN G. PRUITT
NEIL B. McGILLICUDDY
Department of Psychology
State University of New York at Buffalo

A theory is presented about the functions of caucusing (private meetings between the mediator and disputant) in mediation. Empirical results that confirm and extend a number of propositions from the theory are also presented. Two coders, content analyzing both mediator and disputant behavior, observed 51 hearings at the Dispute Settlement Center of Western New York. Results indicated that disputants in caucus sessions employed less direct hostility, provided more information, and proposed more new alternatives than in joint sessions. Mediators, in parallel fashion, were more likely to request information and to challenge the disputants to come up with new alternatives. Mediators also exhibited more freedom to violate the neutrality norms during caucus sessions, giving greater support to the side that originally filed the complaint. These results support the use of caucusing as a route to issue identification and problem solving. However, other results showed that disputants were especially likely to boost their own position and put the other party down during caucus sessions, suggesting that mediators must be wary of what they learn from one party when the other is out of the room.

Mediation can be defined as any effort by a third party to help disputants reach an agreement in a controversy. A major tactic employed by many mediators is "caucusing," that is, meeting separately with each disputant. Kressel (1972) describes caucusing as "the most

frequently discussed tactic of mediation." Mediators differ with respect to whether and how caucusing should be employed (Markowitz and Engram, 1983; Silbey and Merry, 1986). Hence the impact of caucusing on the process and outcome of mediation would appear to be a pressing candidate for research.

The aim of this article is twofold: (a) to present a theoretical discussion of the functions of caucusing that will help understand when caucusing takes place, what is done in caucusing, and what effect these actions have; (b) to present some results from a study that tests certain aspects of this theory.

THE COMPONENTS OF SUCCESSFUL MEDIATION

To understand the functions of caucusing, it is necessary to gain an initial understanding of the components of successful mediation. First and foremost, for mediation to be successful, there is a need for a trusting and empathic relationship between the mediator and each disputant. Thus an important task in the beginning of the hearing, and, to a somewhat lesser degree, throughout the hearing, is to build and maintain rapport with the disputing parties. According to Kolb (1985), there are two elements of rapport building. The mediator must (a) establish his or her credentials and expertise, and (b) establish a positive relationship with the disputants so that they have a sense of his or her "empathic concern" (Kressel, 1972). The first element creates faith that the mediator has the ability to deal with the problem; the second element creates faith that the mediator will invest the necessary effort to deal with the problem.

Rapport contributes to successful mediation in several ways. It facilitates mediator influence over the disputants and makes the disputants more committed to the mediation process (Kelman and Cohen, 1979). Rapport may even contribute to one disputant's taking the needs of the other disputant more seriously, since the mediator, who is trusted, is seen to take these needs seriously (Burton, 1982; de Reuck, 1983).

Once rapport has been established, successful mediation ordinarily goes through four phases: issue identification, generation and evaluation of alternatives, selection of an alternative, and development of an implementation plan (similar stages have been identified by Sheppard, 1984). Movement through these phases is generally forward, though it is

possible to go back to an earlier phase if the decisions made at that time were inadequate.

Issue identification involves understanding a party's position and the interests and assumptions underlying it. Good mediators are often deeply involved in this activity since disputants so frequently lack perspective about what they are demanding and why they are demanding it.

After issue identification, the next step is the generation and evaluation of alternatives. For disputants to generate alternatives, three requirements usually must be met. First, they must be flexible about the means of achieving their interests so that the two sets of interests can be meshed (Pruitt and Rubin, 1986). Second, they must develop a sense of shared responsibility for the problem, instead of considering the other entirely to blame (Sillars, 1981; Syna, 1985). This allows them to shift their attention from who was responsible for past difficulties to the satisfaction of both parties' future interests. Third, they must be able to put up trial balloons without fear of the consequences—that is, to throw out tentative new alternatives without fear of looking weak or of seeming to concede (Pruitt, 1981).

It is best if the two disputants can work together on generating alternatives—that is, if joint problem solving can develop. However, antagonism toward and suspicion of the other party frequently stand in the way of such a development. Alternatively, one party (or both parties) may engage in problem solving separately with the mediator, who hopefully has some grasp of the other party's viewpoint. If all else fails, the responsibility for generating new alternatives may fall exclusively on the shoulders of the mediator.

Having generated and evaluated alternatives, the next step is the selection of an alternative that all parties can accept. When this is accomplished, an additional step, in some but not all hearings, is to develop an implementation plan. Implementation issues must be identified, alternatives generated and evaluated, and a selection made. Hence, when implementation issues arise, it is often necessary to proceed once again through the stages identified above.

ADVANTAGES OF CAUCUSING

Problems often arise in enacting the components of successful mediation discussed in the prior section, especially when the disputants

are antagonistic toward one another. These problems can frequently be resolved in caucus sessions because of four characteristics of these sessions. First, the other party is not present as a stimulus. Hence disputants have less cause to be tense, angry, and defensive and, as a result, should be more flexible and creative. This is especially important when relations are more hostile, as joint sessions tend to become unproductive when hostility levels are high. Second, the other party is not present to hear the disputant. Hence, the mediator should be able to get more information about underlying interests and assumptions, because the other is not there to misinterpret or misuse this information. Also, with the other unable to hear, there should be more willingness to throw out ideas for possible solutions, because the discussion can be kept on a tentative basis without commitment. Third, the other party is not present to see and hear the mediator. This allows the mediator to interact more intimately and warmly with the disputant without appearing partial, which encourages increased rapport and sharing of information. It also allows the mediator to make positive remarks about the other without seeming to curry favor or emboldening the other. Fourth, the other party is not available as a target for the diffusion of responsibility. Hence, the mediator can more effectively challenge the disputant with whom he or she is meeting to take personal responsibility for solving the problem.

Details about the positive benefits of caucusing will be presented in the next two sections, the first dealing with disputant behavior, and the second with mediator behavior.

DISPUTANT BEHAVIOR

There is a consensus in the theoretical literature that caucus (as opposed to joint) sessions are the best place to obtain information about underlying issues, that is, about the interests and assumptions on which the initial demands and the behavior in question are based (Blades, 1984; Evarts et al., 1983; Kolb, 1983; Witty, 1980). The same point has been made about efforts to obtain information about bottom lines, fall-back positions, and priorities among the issues (McGillis, 1981). There are three reasons for this. One is that a close interaction between mediator and disputant may be required to flush out this information. The other party may feel left out if he or she witnesses this interaction. A second is that disputants may be reluctant to provide such information

within earshot of the other for fear of being exploited (e.g., the other may commit himself or herself to the disputant's bottom line) or of being embarrassed, as the other may be critical or spread the stories in the disputant's society. A third is that the motion and defensiveness engendered by having the other in the room may stand in the way of increased self-understanding and understanding of the other. Joint sessions encourage disputants to simply repeat their official positions over and over rather than to explore these positions or listen to one another.

Caucusing can also contribute to problem solving and the development of new alternatives. The reduction in tension and defensiveness occasioned by the other party's departure from the room allows many people to move beyond their initial rigid postures to a more flexible and creative stance (Carnevale et al., 1984). In addition, it is frequently easier for a mediator to focus the responsibility for developing creative alternatives onto a disputant when the other is not in the room as a target for the diffusion of such responsibility. In joint sessions, disputants can more easily tell themselves, "The mediator doesn't mean me, the mediator means the other person because he [or she] is the one to blame for this mess." Finally, it is often much easier for a disputant to mention possible concessions and new alternatives and to respond positively to similar proposals by the mediator in a caucus than in a joint session because the other party is not there to say "I told you so" or to require the disputant to adhere to remarks that were intended to be strictly tentative (Evarts et al., 1983; Haynes, 1981; Kerr, 1954; Kolb, 1983; Markowitz and Engram, 1983; Pruitt, 1981; Witty, 1980).

MEDIATOR BEHAVIOR

Caucus sessions allow the mediator to use tactics that are not especially appropriate in joint sessions. For example, rapport building is aided by sympathetic attention to and empathic concern for a disputant. Sometimes it is necessary to build more rapport with one disputant (e.g., the party who feels more aggrieved or more suspicious) than with the other. This is difficult in a joint session, because the other may interpret unequal sympathetic attention as an official recognition of the opponent's position, leading to a perception of the mediator as biased. Hence it is often better to pursue this in caucus.

In addition to the issue of unequal sympathetic attention, even equal sympathetic attention that is provided to both parties in joint session

might result in both parties' selectively perceiving that the mediator was biased in favor of the other side (see Vallone et al., 1985).

It is sometimes valuable for the mediator to make positive remarks about the other party and his or her position so as to encourage a party to have concern about the other's needs. Such statements are more appropriate in caucus sessions, as otherwise the target of these remarks may regard them as attempts to curry favor with the other party and hence as signs of mediator bias (Pruitt, 1981). Caucusing also allows the mediator to question the assumptions and perspectives of the disputant.

Caucus sessions are probably also the best place to challenge disputants to devise new alternatives. One reason is that disputants are likely to be less emotional and defensive and hence more flexible in caucus than in joint sessions. The other is that, with only one disputant in the room, there is less likelihood of a diffusion of responsibility for thinking up new ideas.

Caucus sessions also allow the mediator to encourage a disputant to make concession or to accept a particular alternative without embolden-ing the other party or risking the mediator's image of impartiality (Bercovich, 1984; Blades, 1984; Kolb, 1983). Such a strategy in joint session would tend to cause resentment in the party who is being pressed and rigidity in the other party because the mediator seemed to be legitimizing his or her position. It is particularly important for mediators to employ caucus sessions when it appears that one party must make most of the concessions, because otherwise their image of impartiality will be questioned (Shapiro et al., 1985).

Mediators also have more influence in caucus sessions. This is because they can capitalize on the interpersonal bond and press for concessions without emboldening the other party or appearing to be acting in consort with the other.

Caucus sessions also allow the mediator to pass ideas from one disputant to the other without revealing the source of these ideas. This makes it possible for a tentative concession to be explored with the other party without emboldening that person. If the other does not recipro-cate, the concession may be withdrawn.

Finally, caucusing provides mediators with an opportunity to empower a weak disputant. When one side is weaker than the other, there is danger of a one-sided agreement that will benefit neither side in the long run. Mediators can avoid this outcome by encouraging the weaker side to stick to its guns. But this can be done only in caucus, because otherwise it would alienate the stronger side.

DISADVANTAGES OF CAUCUSING

While many benefits of caucusing can be enumerated, the picture is not entirely rosy. Caucusing can backfire. One problem is that disputant statements in caucus sessions are more likely to be inaccurate than those in joint sessions, because the opponent is not there to correct them. A disputant can make wild accusations or employ specious arguments without fear of rebuttal from the other side. A second problem is that the caucusee may coopt the mediator, unduly impressing the latter with the strength of his or her resolve or the correctness of his or her position. In other words, the caucusee, like the mediator, can have more influence than in a joint session.

A third problem is that the party who is not in caucus may be meeting with friends, relatives, or other supporters, or developing new arguments, resulting in the solidification of this position. Finally, caucusing may interrupt a joint problem-solving discussion between the disputants. Hence mediators must be careful not to caucus prematurely.

METHOD

The data were gathered as part of an experimental study that investigated the impact of three types of third-party intervention on the process of mediation at the Dispute Settlement Center (DSC) of Western New York, a division of the Better Business Bureau Foundation of Western New York, Inc. This data enables many, but not all, of the above hypotheses to be tested.

The DSC, under contract with the Unified Courts System of the State of New York, provides low-cost, confidential, and informal hearings as a practical alternative for dispute resolution. The service is made available for individuals who have some type of ongoing relationship. This includes relations between neighbors, friends, family members, housemates, workmates, and so forth. The DSC provides a trained volunteer mediator to assist the parties to work out a settlement. The parties are encouraged to go beyond the original charge through a discussion of the underlying issues at the root of the problem. High levels of compliance with the decisions have been reported (Syna et al., 1983).

Most cases are referred from Buffalo City Court and Buffalo Family Court. The center also is available for walk-ins. The typical case in our sample came as a result of one party (the complainant) going to Buffalo

City Court and attempting to take out a warrant against a second party (the respondent). Both parties then appeared at a short hearing during which a prewarrant officer decided on the appropriateness of the case for the DSC. Disputants chose whether or not to participate in the DSC program.

The disputes in our sample concerned such matters as financial restitution, complaints about behavior, return of property, and visitation schedules. The relationships included friends and ex-friends, ex-lovers, neighbors, family, coworkers, and, in two cases, strangers.

Once the disputants had agreed to mediation, the DSC scheduled the hearing and then assigned a mediator who could appear at the time. Mediators had an opportunity to refuse to participate in the study at the time they were called, and two did so. At the beginning of the hearing, the disputants were also given an opportunity to refuse to participate in the study, which they did in three hearings.

The mediator began the hearing by introducing him- or herself to the parties and explaining the ground rules. The hearing was then conducted. The typical hearing began with a venting stage, in which each side told his or her story. The mediator then worked at identifying the issues. Next, the mediator tried to develop a solution or to persuade the parties to do so. When an agreement was reached, a consent decision was written. The hearings varied in length from 20 to 160 minutes.

The volunteer mediators ranged in experience from three months to six years. All had participated in a 40-hour training session prior to becoming a mediator. As part of their training, they had been observed in at least one hearing and judged to be competent. They were heterogeneous in race, ethnic background, and sex.

RECORDING PROCEDURE

During the hearing, the two observers sat in a corner of the room as quietly and inconspicuously as possible. They coded the statements of the mediator and disputants into categories on an "OS 3" electronic event recorder with two response pads. Coding took place in both joint sessions and caucus sessions. Altogether there were 26 categories for the disputants (e.g., asks a hostile question, proposes a new idea) and 28 for the mediator (e.g., requests a proposal, threatens to terminate the session). In the 15 reliability hearings, both observers coded for the same set of categories (the mediator was observed in eight hearings and the disputants in seven). For the 36 experimental hearings, the observers

coded separate categories, one being assigned to the mediator and the other to the disputants.

Of the 26 disputant codes, 5 were rare and 6 others were not related to these hypotheses. Therefore, only 15 codes will be discussed. Of the 28 mediator codes, 2 were very rare. All of the remaining 26 codes were related to the hypotheses. One subset of 5 codes concerned behavior by the mediator directed toward the complainant, and another parallel subset of 5 codes concerned behavior directed toward the disputant. For many of the analyses, these codes were summed across the target (complainant and respondent), resulting in 5 codes rather than 10. Hence much of the analysis involved 21 mediator codes.

SUBJECTS

The original sample included 51 mediation hearings, 2 of which had to be excluded from the analysis because they were so long that they exceeded the memory of the recording apparatus, resulting in the loss of some data. Each hearing included one mediator, at least one complainant ($M = 1.22$), and at least one respondent ($M = 1.22$). The primary disputant on each side was an adult. The disputants were also heterogeneous in race, ethnic background, and sex. Four hearings included one or two attorneys, and five included one or more witnesses who were brought in to testify for a brief time.

Some of these 49 hearings, however, did not employ any caucuses, the use of which depended on the judgment of the mediator about its need or potential benefit. At least one caucus was employed in 32 of the remaining 49 hearings. (In every hearing that employed a caucus, at least one caucus was held with each side.) These caucuses typically included the mediator and all members of one side (including an attorney, if one was involved) but did not include any witnesses. In these 32 hearings, 65% of the statements were made in joint session and 35% in caucus. The percentage of statements made in caucus for these hearings ranged from as little as 12% to as much as 77%.

Since most of our questions concerned how behavior in the caucus sessions differed from that in the joint sessions, it was appropriate to compare only those hearings that employed both joint and caucus sessions. Based on this requirement, our final sample for most of the analyses included two overlapping subsets of the 32 hearings that employed caucuses: 28 hearings with complete data on the mediator (5 from the reliability set and 23 from the experimental set) and 27 with

complete data on the disputants (4 from the reliability set and 23 from the experimental set).

DEPENDENT MEASURES

The on-site coding of verbal statements was done as follows: The coding unit was everything said by one party (whether disputant or mediator) between statements of another party or parties, provided that the statement did not exceed 30 seconds. If a party spoke for more than 30 seconds, a unit was initiated, and so on for every 30-second interval. A unit could receive several different codes if it contained the appropriate themes. However, a single unit could not receive the same code twice.

Several questionnaire measures were included at the end of the hearing, largely for the purpose of the experimental study. Of interest in this context are responses to an open-ended question asking those mediators who caucused what they had hoped to achieve in the caucus sessions.

While it would be useful to have an outcome measure to use as a criterion variable, attempts to measure satisfaction and compliance were hampered by the refusal of many disputants to complete questionnaires and the difficulty of contacting disputants several months later to interview them about compliance with the agreement. Alternative measures of outcome have been developed for use in a follow-up study.

STATISTICAL ANALYSIS

Of primary interest was the difference between participants' behavior in joint versus caucus sessions. On the assumption that these differences might be moderated by the phase of the mediation process, the hearing was divided into three equal time periods, which we call "stages."

This procedure might seem to imply the use of a 2 (joint vs. caucus) × 3 (early, middle, late) within-group factorial analysis of variance. The design was rejected, however, since there were very different N's in each cell owing to the fact that caucuses came at different points in different sessions. For example, only 13 hearings employed a caucus during the first stage, and only five hearings employed both a joint and a caucus session in each of the three stages. Hence, the following procedure was used for the primary analysis. All hearings (28 for mediator behavior, 27

for disputant behavior) were included in tests of the main effects of joint versus caucus and of stages (the latter of which will not be reported in this article). Tests of the simple effects of joint versus caucus sessions at each stage were made using only those hearings that had both joint and caucus sessions during that stage.

Relative frequency scores were used throughout the analysis. Mediator scores were constructed by dividing the frequency of each code in each cell by the number of mediator units in that cell. Similarly, disputant scores were constructed by dividing the frequency of each code in each cell (summed across disputants) by the number of disputant units in that cell (summed across disputants). Complainant and respondent scores were combined for most analyses because the recording methods did not permit distinguishing between these two parties in joint sessions. Since a distinction could be made in caucus, additional analyses were done to compare complainant versus respondent statements in caucus and to compare mediator statements directed to the complainant with those directed to the respondent.

The conditions that led to the use of caucusing were also of interest. This analysis employed the 34 experimental hearings for which there was complete data on both the disputants and the mediator. Comparisons were made between the 23 hearings that employed caucusing and the 11 that did not. Only data from the first stage were employed, as it seemed reasonable that the events of this stage would largely determine whether the mediator would choose to call a caucus.

INTERCODER RELIABILITY

Pearson product-moment correlation coefficients were calculated for each of the mediator and disputant codes using data from the reliability hearings. A total of 24 of the 28 mediator codes and 13 of the 15 disputant codes reached a criterion level of .70. Analyses were conducted on all of the codes. Findings associated with the unreliable codes will be so identified and should be considered highly tentative.

RESULTS AND DISCUSSION

The primary aim of the current article is to explore possible mechanisms whereby caucusing may have an impact on disputant and

mediator behavior. To ensure that findings on the 21 mediator codes and 15 disputant categories did not simply result from chance, four multivariate analyses were conducted, with joint versus caucus as the within-group factor.

The first analysis was conducted on the entire group of 21 mediator codes. The results indicated that mediators behaved differently in caucus than in joint sessions, $F(21,7) = 29.26$, $p < .001$.

The 15 disputant codes were divided into three sets of 5 codes each. The first set included measures of problem-solving behavior (giving information, giving priority information, requesting the other's reaction to a proposal, offering a new proposal, making a concession). The results indicated a trend in the data, $F(5,22) = 2.234$, $p < .09$. The second set included measures of contentious behavior (comparisons between self and other, self-enhancement, threats, heavy commitments, and miscellaneous persuasive arguments). The results indicated a nonsignificant trend in the data, $F(5,22) = 2.025$, $p < .12$. The third set included measures of hostility (hostile questions, complaints about the other's character, complaints about the other's behavior, sarcasm, and swearing and angry displays). The results indicated a difference in hostility between the joint and the caucus sessions, $F(5,22) = 3.628$, $p < .02$.

REDUCED DIRECT HOSTILITY

Results from two disputant codes indicate that direct emotion-laden hostility was reduced in caucus sessions. Swearing and angry displays, which included swearing at or about the other disputant, screaming and yelling, and hitting the table, occurred more frequently (per coding unit) during joint sessions (M = .0308) than during caucus sessions (M = .0114), $t(26) = 2.04$, $p < .06$. Hostile questions, which included any question asked in a hostile tone, also occurred more frequently during joint sessions (M = .0142) than during caucus sessions (M = .0017), $t(26) = 3.32$, $p < .01$. These findings suggest, as hypothesized earlier, that emotional tension goes down when the other party leaves the room.

Two mediator codes associated with social control provide additional evidence of reduced direct hostility in caucus sessions. Negative evaluation of the parties' behavior occurred more frequently during joint sessions (M = .0336) than during caucus sessions (M = .0039), $t(27) = 3.34$, $p < .010$. Efforts to reduce the level of hostility also occurred more frequently during joint sessions (M = .0038) than during caucus

sessions (M = .0004), $t(27)$ = 1.95, $p < .07$. These two tactics were used primarily during the first stage of the hearing, when the discussion was usually the most heated.

These findings provide evidence that caucus sessions are an effective way of reducing direct emotion-laden hostility. Several mediators, in fact, reported that hostility control was one of their major reasons for caucusing. One hearing that included very high levels of hostility employed caucusing almost exclusively (77%), presumably so as to avoid this direct hostility.

INCREASED INDIRECT HOSTILITY AND PERSUASIVE ARGUMENTS

Though direct hostility was diminished in caucus sessions, "indirect" types of hostility were on the increase. The disputants tried harder to impress the mediator with the wrongness of their opponent's cause. Behavioral putdowns, in which a disputant criticized the other party's behavior (e.g., "He cheated me last month."), occurred more frequently during caucus sessions (M = .1460) than during joint sessions (M = .1120), $t(26)$ = 1.71, $p < .10$. Character putdowns, in which disputants criticized some character trait of the other party (e.g., "He is a cheater."), also occurred more frequently during caucus sessions (M = .0197) than during joint sessions (M = .0103), $t(26)$ = 2.52, $p < .05$.

Two contentious codes demonstrated a similar pattern to these two hostile codes. Self-enhancement, in which disputants tried to put themselves in a better light by emphasizing their own good qualities, occurred more frequently during caucus sessions (M = .0666) than during joint sessions (M = .0322), $t(26)$ = 3.68, $p < .001$. Miscellaneous persuasive arguments, which included other statements used to support one's position, were also more frequent during caucus sessions (M = .2990) than during joint sessions (M = .2139), $t(26)$ = 3.82, $p < .001$. (This final result is largely due to a decrease in usage over time in joint sessions, compared to a constant use in caucus sessions.)

In short, the disputants used the caucus sessions disproportionately to praise themselves and criticize their adversaries. This can probably be explained by the fact that the other party was not present in caucus sessions to make disclaimers. Though we cannot be sure from the data, it seems likely that many of these statements were less than truthful, as suggested in the earlier theoretical discussion. That such statements could nevertheless have an impact on the mediator is suggested by the

finding that mediators tend to recommend solutions that favor the side that has been most vigorous in presenting his or her position (Harnett and Wall, 1983).

INCREASED SOLICITATION AND
GENERATION OF NEW ALTERNATIVES

As anticipated in the earlier theoretical discussion, the mediator asked the parties to present a new alternative more often during caucus sessions (M = .1212) than during joint sessions (M = .0744), $t(27) = 2.47$, $p < .05$. Moreover, the mediators were granted what they requested in that the disputants offered more new alternatives, (or new variants of an old alternative) during caucus sessions (M = .0558) than during joint sessions (M = .0322), $t(26) = 2.43$, $p < .05$. This supports the hypothesis that there is less rigidity and more capacity for creative thinking in caucus than in joint sessions. A possible explanation for this effect is that emotion and defensiveness are reduced in caucus sessions due to the absence of the other party. In addition, it may be easier for a disputant to put forward a tentative new idea in caucus than in joint session because the other is not there to view it as a sign of weakness.

ASKING AND GIVING INFORMATION

The mediator asking for information code, which included a request by the mediator for any kind of information from one or both parties, declined rapidly over time in joint sessions, but remained fairly constant in caucus sessions (see Table 1 for means). In the first stage, asking for information occurred more frequently during joint than during caucus sessions, $t(10) = 2.64$, $p < .05$. In the last stage, however, asking for information occurred more frequently during caucus than during joint sessions, $t(23) = 2.56$, $p < .05$. The differences in pattern over time suggest that the information sought in caucus is probably very different from that sought in joint sessions. In support of this interpretation, several mediators reported that information gathering was an important reason for their use of caucusing, and that they believed that more information about disputants' bottom lines could be obtained in caucus as opposed to joint sessions. The decline during joint sessions presumably reflects a shift in goals away from defining the issues. The steady use in

TABLE 1
Information Exchange

Mediator Code
  Asking for information (proportion of units)

	Joint	Caucus
First stage	.3382	.1988
Middle stage	.2011	.1681
Final stage	.0522	.1421

Disputant Code
  Giving information (proportion of units)

	Joint	Caucus
First stage	.5816	.5507
Middle stage	.4960	.5673
Final stage	.3261	.5433

caucus sessions may indicate that an important element of caucuses at all stages is a request for information relating to the nature of the impasse in the joint session that occasioned calling the caucus. These interpretations are speculative inasmuch as our measure of information requested did not permit a fine-grained analysis of the kinds of information sought.

The "disputant giving information" code, assigned to disputant statements that provided information of any kind, was a frequently used category. The pattern for this measure was similar to that for the "mediator requesting information," in that giving information remained constant in caucus but declined in joint sessions (see Table 1). In this case, however, the means were not different during the first stage, but became increasingly discrepant. This resulted in a main effect, such that giving information occurred more frequently during caucus sessions (M = .5458) than during joint sessions (M = .4783), $t(26) = 2.39$, $p < .03$, a result that was mainly apparent in the last stage, $t(21) = 4.26$, $p < .001$. Again, the on-site coding methodology did not permit a fine-grained analysis of the kinds of information being provided, though it seems likely that the information in caucus more often concerned interests, principles, bottom lines, and priorities among the issues.

THE EVOLUTION OF PROBLEM SOLVING

Several codes provide evidence that certain forms of problem solving tended to originate in middle-stage caucus sessions and to be carried over into the joint session in the final stage. The analysis employed here was a 2×2 within-groups factorial analysis of variance, with the first factor being joint versus caucus session and the second factor being middle versus final stage. This analysis used only the 15 hearings that employed a joint and a caucus session in both the middle and the final stages, resulting in a limited number of degrees of freedom.

Requests for reaction, in which a disputant asked the mediator or the other disputant to react to an alternative, occurred most frequently in middle-stage caucus sessions and final-stage joint sessions, $F(1,14) =$ 4.60, $p < .05$ (see Table 2). Such requests are an element of problem solving. A second disputant problem-solving code, offering new alternatives, showed a similar pattern, but the interaction did not approach significance.

A similar pattern can be seen in the data for mediators offering new alternatives. Mediators tended to offer new alternatives in caucus sessions during the middle stage and then in joint sessions during the final stage, $F(1,14) = 2.87$, $p < .12$ (see Table 2).

These findings suggest that new ideas tend to be tested first in caucus to be sure they are acceptable to the individual parties and then introduced into the joint session. The new alternatives suggested by the mediator in the joint session were presumably a combination of the individual alternatives discussed with the two parties in the caucus sessions.

All of this suggests that, as indicated earlier, caucusing can be a route into problem solving. Caucusing permits the exploration of new ideas, generated by both the mediator and the disputants, which can then be more fully developed in joint session.

MEDIATOR DIFFERENTIAL SUPPORT FOR DISPUTANTS

Two related mediator codes, positive evaluation of the complainant's position and positive evaluation of the respondent's position, were analyzed by means of a 2 (joint vs. caucus session) by 2 (complainant vs. respondent) analysis of variance (see Table 3 for means). The interaction between these factors was nearly significant, $F(1,27) = 4.04$, $p < .06$.

TABLE 2
Problem Solving

		Joint Sessions	*Caucus Sessions*
Mediator Code			
New proposals	Middle stage	.0195	.0400
(p < .12)	Final stage	.0600	.0217
Disputant Code			
Request reaction	Middle stage	.0084	.0097
(p < .05)	Final stage	.0143	.0042

Positive evaluation of the two parties was equal in joint session. In caucus, however, the complainant was evaluated positively more often than the respondent, regardless of which disputant was present in the caucus. This suggests, as proposed earlier, that it is safer for mediators to side with one of the parties in caucus than in joint session. The problem with doing so in joint session is that it becomes obvious that the mediator is taking such a stance, which is likely to hurt his or her empathic relationship with the party not favored.

Why would the mediator tend to favor the complainant over the respondent? There are several possible explanations. One is that the mediator may be better able to empathize and identify with the complainant, who usually claims to be a victim, than with the respondent, who is alleged to be the perpetrator. A second explanation is that the structure may dictate this stance, in that the mediator must lean toward the complainant (who is usually trying to effect some change) in order to balance the force of the status quo, since the latter usually favors the respondent. A third possible explanation is that the complainant has a stronger case than the respondent, causing the mediator to come down on his or her side more frequently.

TACTICS MORE COMMON IN JOINT SESSIONS

Three types of mediator behavior—procedural suggestions, threats, and urging agreement—were more common in joint than in caucus sessions.

TABLE 3
Mediator Differential Support for Disputants

	Joint Sessions	Caucus Sessions
Mediator Codes		
Positive evaluation of complainant's position	.0071	.0140
Positive evaluation of respondent's position	.0065	.0045
		p < .06

Procedural suggestions. These suggestions were coded in two ways: those that were related to the agenda (concerning how the agenda would be ordered, what issue would be discussed next, etc.) and those that were not related to the agenda (concerning who would speak next, when a caucus would begin or end, who would sign the legal document first, etc.). Note that intercoder reliability did not meet the criterion for either of these codes. With that in mind, the results demonstrated that both types of procedural suggestions occurred more frequently during joint sessions than during caucus sessions (agenda/joint M = .0112, agenda/caucus M = .0017, $t(27) = 2.60$, $p < .05$; nonagenda/joint M = .0976, nonagenda/caucus M = .0628, $t(27) = 2.99$, $p < .01$). These findings probably result from the fact that procedural suggestions are devices for coordinating the two disputants' activities. It is logical that coordination suggestions be introduced in joint sessions because they are more likely to work if each party has reason to believe that the other party has heard the same suggestion.

Mediator threats to withdraw. The mediator sometimes threatened to end the hearing as a device to encourage concessions. These threats occurred more frequently during joint sessions (M = .0115) than during caucus sessions (M = .0021), $t(27) = 1.86$, $p < .10$, perhaps because such statements would weaken rapport if one believed the mediator was threatening him or her without threatening the other party.

Urging agreement. Mediator statements urging the parties to come to an agreement occurred more frequently during joint sessions (M =

TABLE 4
Conditions That Lead to the Use of Caucusing

	$p <$	*Joint Only*	*Joint w/Caucus*
Mediator Codes			
Positive evaluation of position	.08	.060	.004
Negative evaluation of behavior	.08	.013	.088
Disputant Codes			
Request reaction	.06	.017	.002
New proposals	.02	.054	.013
Misc. persuasive arguments	.01	.142	.283

.0058) than during caucus sessions (M = .0024), $t(27) = 1.91$, $p < .10$, perhaps because the same goals that led to the mediator to urge the parties to reach an agreement in joint sessions instead led the mediator to encourage the development of new alternatives in caucus sessions.

CONDITIONS LEADING TO THE USE OF CAUCUSING

Five codes, as measured during the first stage, distinguished the cases that employed a caucus from those that did not (see Table 4 for means). In cases that employed a caucus, disputants used more persuasive arguments, made fewer requests for reaction to an alternative, and generated fewer new alternatives. Mediators employed more negative evaluations of the parties' behavior and less positive evaluations of their positions during these cases. These findings suggest that mediators tend to call caucuses when disputants are taking a contentious, as opposed to problem-solving, approach.

CONCLUSIONS

The use of caucus sessions is an important tool in the mediation process. To the extent that we can increase our understanding of the potential benefits and costs of caucusing, we can increase the usefulness of this form of dispute resolution.

The results of the study described in the latter half of this article support and help to elaborate a number of hypotheses mentioned in the

earlier theoretical discussion: (1) Direct, emotion-laden hostility is less prominent in caucus than in joint sessions. (2) Disputants make more effort in caucus than in joint sessions to advance their interests by means of self-enhancing statements and statements that put the other down. This is compatible with the hypothesis that the mediator receives more false information in caucus than in joint sessions. (3) Disputants are more likely to propose new alternatives in caucus than in joint sessions. This may reflect a lower level of psychological tension or a reduced danger of appearing weak in caucus as opposed to joint sessions. (4) Mediators are more likely to request new alternatives from the disputants in caucus than in joint sessions. This may reflect an awareness that disputants are more capable of generating such alternatives or a perception that it will be easier to persuade the disputants to take responsibility for doing so during caucus. (5) In the final stage of mediation, mediators tend to request and disputants to give more information during caucus than during joint sessions. (6) Caucusing allows mediators to take sides with one party in order to move the process along. Thus it appears that caucusing somewhat relieves the third party of the requirement of being strictly neutral between the two parties. (7) Problem solving tends to start in caucus sessions and then migrate to joint sessions. (8) Mediators make more procedural suggestions and exert more pressure to reach agreement in joint than in caucus sessions.

These findings suggest that mediators should and do find caucus sessions useful for pushing the process of dispute resolution along when little progress is being made in joint session. However, since disputants tend to be especially critical of the adversary when the latter is out of earshot, mediators must be wary of information obtained during caucus sessions.

REFERENCES

BERCOVITCH, J. (1984) Social Conflicts and Third Parties: Strategies of Conflict Resolution. Boulder, CO: Westview.
BLADES, J. (1984) "Mediation: an old art revisited." Mediation Q. 3: 59-95.
BURTON, J. W. (1982) Dear Survivors: Planning After Nuclear Holocaust. London: Frances Printer.

CARNEVALE, P., S. BECKMAN, and S. BONA (1984) "Cognitive elements of cooperation and competition: categorization and functional fixedness." Submitted to the annual meeting of the American Psychological Association, Toronto.

De REUCK, A.V.S. (1983) "A theory of conflict resolution by problem solving." Man, Environment, Space and Time 3: 1-14.

EVARTS, W. R., J. L. GREENSTONE, G. J. KIRKPATRICK, and S. C. LEVITON (1983) Winning Through Accommodation: The Mediator's Handbook. Dubuque, IA: Kendall/Hunt.

HARNETT, D. L. and J. A. WALL, Jr. (1983) "Aspiration/competitive effects on the mediation of bargaining," pp. 8-21 in R. Tietz (ed.) Aspiration Levels in Bargaining and Economic Decision Making. Berlin: Springer-Verlag.

HAYNES, J. M. (1981) Divorce Mediation: A Practical Guide for Therapists. New York: Springer.

KELMAN, H. C. and S. P. COHEN (1979) "Reduction of international conflict: an interactional approach," pp. 288-303 in W. G. Austin and S. Worchel (eds.) The Social Psychology of Intergroup Relations. Monterey, CA: Brooks/Cole.

KERR, C. (1954) "Industrial conflict and its mediation." Amer. J. of Sociology 60: 230-245.

KOLB, D. (1983) The Mediators. Cambridge: MIT Press.

KOLB, D. (1985) "To be a mediator: expressive tactics in mediation." J. of Social Issues 41: 11-26.

KRESSEL, K. (1972) Labor Mediation: An Exploratory Survey. Albany, NY: Association of Labor Mediation Agencies.

MARKOWITZ, J. R. and P. S. ENGRAM (1983) "Mediation in labor disputes and divorces: a comparative analysis." Mediation Q. 2: 67-78.

McGILLIS, D. (1981) "Conflict resolution outside the courts," pp. 243-262 in L. Bickman (ed.) Applied Social Psychology Annual, Vol. 2. Beverly Hills, CA: Sage.

PRUITT, D. G. (1981) Negotiation Behavior. New York: Academic Press.

PRUITT, D. G., and J. Z. RUBIN (1986) Social Conflict: Escalation, Stalemate and Settlement. New York: Random House.

SHAPIRO, D., R. DRIEGHE, and J. BRETT (1985) "Mediator behavior and the outcome of mediation." J. of Social Issues 41: 101-114.

SHEPPARD, B. H. (1984) "Third party conflict intervention: a procedural framework." Research in Organizational Behavior 6: 141-190.

SILBEY, S. S. and S. E. MERRY (1986) "Mediator settlement strategies." Law and Policy 8: 7-32.

SILLARS, A. L. (1981) "Attributions and interpersonal conflict resolution," pp. 279-305 in J. H. Harvey et al. (eds.) New Directions in Attribution Research, Vol. 3. Hillsdale, NJ: Lawrence Erlbaum.

SYNA, H. (1985) "Conflict resolution strategies and blaming patterns in conflicts of intimate couples." Presented at the annual meeting of the Eastern Psychological Association, Boston.

SYNA, H., M. VAN SLYCK, N. McGILLICUDDY, T. NOCHAJSKI, and D. G. PRUITT (1983) "Effectiveness of community mediation for conflict resolution: an evaluation." Presented at the annual meeting of the American Psychological Association, Anaheim, CA.

VALLONE, R. P., L. ROSS, and M. R. LEPPER (1985) "The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the Beirut massacre." J. of Personality and Social Psychology 49: 577-585.

WITTY, C. J. (1980) Mediation and Society: Conflict Management in Lebanon. New York: Academic Press.

The strategic use of caucus to facilitate parties' trust in mediators

Jean Poitras

*Department of Human Resources Management, HEC Montréal,
Montréal, Canada*

Abstract

Purpose – Mediators' impartiality and empathy are two classical factors in the parties' trust in mediators. However, mediators are often torn between being impartial and being empathetic. The aim of this paper is to explore this empirically.

Design/methodology/approach – This study empirically tests the strategic use of caucus to improve the interaction between impartiality and empathy by splitting them into two phases: impartiality in joint sessions and empathy in caucus.

Findings – The strategy did create significant synergy between impartiality and empathy with the main impact of reducing the time needed to reach an agreement.

Research limitations/implications – All research data come from workplace mediation and from the same organization. Although it can be reasonably postulated that the results can be generalized to other mediation settings, this remains to be proven.

Practical implications – When mediators use the trust caucus strategy, impartiality and empathy work better together and parties put more weight on empathy than on impartiality. While the use of the trust caucus does not increase the likelihood of reaching agreement, it does significantly decrease the time needed to conclude an agreement.

Originality/value – The study uses a quasi-experimental design to test its hypothesis. Furthermore, the study uses real mediation cases.

Keywords Mediation, Impartiality, Empathy, Trust, Caucus, Negotiating

Paper type Research paper

For a mediation to be successful, it is legitimate to think that trust in the mediator is required. First, trust is the point of departure from which the mediator can bring to bear his or her expertise. In fact, studies have shown that parties who perceived a mediator bias toward the other party were much less accepting of the mediator's actions (Welton and Pruitt, 1987). Second, trust is also important because it determines whether a party will feel comfortable with openly stating his or her needs in the context of mediation (McCarthy, 1985). As a result, parties who distrust the mediator may not provide access to important information (Butler, 1999).

Establishing a mediator's trustworthiness is a complex process, mainly because of triangulation (Augsburger, 1992). Mediators are by definition "go betweens" that link two disputing parties. Their role can be limited to simply purveying offers from one side to the other, or can be much more complex, as in restoring relationships. In all cases, mediators must have strong rapport with both parties to be efficient. As a result, mediators have to build trust with both parties. For a mediator, a difficult challenge is often to build trust with one party without losing the other party's trust in doing so.

This paper studies how parties' trust in the mediator is built, and how the strategic use of caucus can help the mediator build rapport with both parties more efficiently. More specifically, we measure the effect of a trust caucus strategy that was first taught to mediators in a training session. To assess our hypotheses, we use a quasi-experimental design with a large sample of real mediation cases, focusing on the parties' point-of-view.

Trust, impartiality and empathy

Conditions that lead to trust have been studied frequently in the literature. Some authors identified a single trustee characteristic that is responsible for trust (Strickland, 1958), whereas other others define as many as ten characteristics (Butler, 1991). The Mayer *et al.* (1995) integrative model of trust proposed three key characteristics of a trustee: ability, benevolence and integrity.

In mediation, ability is the group of mediator skills, competencies and characteristics that make parties' feel that the mediator can move the mediation process forward. Benevolence is the extent to which a mediator is believed to do good to parties. According to Mishra (1996), caring is a key component of benevolence. Consequently, mediators demonstrate benevolence when they show some kind of empathy toward parties. Integrity is related to the parties' perception that the mediator adheres to a set of principles that parties find acceptable. According to Butler (1991), fairness is a key component of integrity. Thus mediators show integrity when they demonstrate fairness, which is often associated with impartiality in mediation (Moore, 2003).

While all three factors are important, we will focus more specifically on impartiality and empathy in this paper because these two ingredients of trustworthiness are often difficult to combine. By empathizing with one party, the mediator risks being perceived as partial by the other (Goldberg, 2005). As a result, mediators are often torn between being impartial and being empathetic. How can one be empathetic without appearing to take sides?

Mediators' impartiality and empathy

Mediators' neutrality and impartiality are frequently cited components of a mediator's trustworthiness (Albin, 1993). In the mediation literature, there are three basic hypotheses that could explain how a party comes to perceive a mediator as impartial. First, justice and impartial management are similar concepts that have been identified in the literature (Zaheer *et al.*, 1998). In mediation, this refers to activities such as informing parties without showing any bias or even maintaining procedural equity in managing the process (Moore, 2003). Second, individuals are more likely to trust others if they believe the other person has nothing to gain from untrustworthy behavior (Deutsch, 1958). It is therefore important for mediators to show the parties that they have no interest in favoring one party over another. Furthermore, mediators have to show that they will not be rewarded as a function of the conflict's outcome. Third,

impartiality may also encompass the perception that the mediator is concerned with finding a just solution (Augsburger, 1992). This is similar to the concept of "balanced partiality" (Wehr and Lederach, 1991). With this concept of neutrality, a mediator has a bias toward finding a win-win solution (i.e. a balanced outcome in which every party gains something). Because of the importance of impartiality, we postulate that when mediators demonstrate impartiality, they gain parties' trust:

H1. Greater mediators impartiality is positively related to parties' trust.

It is almost a truism that mediator should be empathetic and impartial. For example, Christopher Honeyman (1988) argues that one of the five elements for successful mediator is empathy. For McKnight and Chervany (2001), the idea of empathy involves two key elements: goodwill and receptiveness. First, Pruitt and Lewis (1983) notes that, to feel at ease, parties must have the impression that the mediator has a positive image of them. Parties must somehow feel the mediator cares about them. Second, Augsburger (1992) says that parties must feel that the mediator cares about their concerns. It is only when parties feel a mediator is receptive to their interests that they will feel comfortable with the mediator's interventions such as querying for information or suggesting concessions. More specifically, Zubeck *et al.* (1992) found that empathy was important for establishing trust in mediation. Consequently, we postulate that mediators' empathy will be positively linked with parties' trust:

H2. Greater mediator empathy is positively related to parties' trust.

Challenge of rapport building in mediation
A key step in the mediation process occurs when the mediator builds a rapport with each party (i.e. when a mediator establishes his or her friendliness and trustworthiness). During this step, the mediator sets up a positive relationship with the disputants so that they have a sense of his or her "empathic concern" (Kressel, 1972). A good rapport contributes to successful mediation by facilitating mediator influence over the parties and makes them more committed to the mediation process (Kelman and Cohen, 1979).

However, building rapport in the context of a mediation can be quite tricky. According to Stephen Goldberg (2005), the mediator must tread carefully in his or her rapport building activities to avoid appearing to be sympathetic to the other. Moreover, according to Vallone *et al.* (1985), even equal sympathetic attention that is provided to both parties might result in both parties' selectively perceiving that the mediator was biased in favor of the other side. Consequently, while impartiality and empathy are core elements of parties' trust in their mediator, these two elements might not be compatible. We therefore postulate that they have a negative interaction in regard to establishing a mediators' trustworthiness:

H3. Mediators impartiality and empathy interact negatively to hinder trust.

Caucus as a tool to build rapport
The proposed negative interaction of impartiality and empathy is not unsolvable. In fact, Deborah Kolb (1985) proposed that using caucus with one party is an excellent strategy to promote rapport building free from scrutiny by the other party. According to Welton *et al.* (1988), because the other party is not present to see and hear the mediator, he or she can interact more intimately and warmly with the disputant without appearing partial.

This encourages increased rapport and sharing of information. This is congruent with the caveat that mediator should be empathic but should not show partiality in joint session (Desivilya and Rotem, 2008). As a result, we propose the systematic use of a caucus to split mediator demonstration of impartiality and empathy. We call the use of caucus to facilitate rapport building the trust caucus strategy.

More specifically, the trust caucus strategy proposes demonstrating impartiality in joint meetings while exhibiting empathy in caucus. The basic idea is that in private caucus, the mediator may show more empathy to the parties (Sherman, 2003). The strategy can be broadly described in two steps. First, in the joint session, they must learn to be more oriented toward impartial process management (even if it makes them appear colder). For example, they should stress parties' equal opportunity to voice their concerns. Second, in the caucus session, they must relate to parties' preoccupations (even if it makes them appear less impartial). For example, the mediator could summarize a party's concerns and say he or she understands that these points are important to the party.

The trust caucus strategy is proposed as a mediation loop within the mediation process. We are not suggesting that after the first joint session the mediator should proceed only with caucus. Instead, we are proposing that the mediator use a quick caucus to build trust early in the first joint session. After that short caucus, the mediator can proceed with joint meeting or further caucus depending on what is needed to move forward the resolution of the conflict.

It is interesting to note that the proposed strategy is in line with the finding of Welton *et al.* (1988) that showed that mediators exhibited more freedom to "violate" the neutrality norms during caucus session. With the trust caucus strategy, we are proposing using this natural mediator propensity as a systematic strategy to build trust. More specifically, we postulate that such a trust caucus strategy will create a positive interaction between impartiality and empathy:

H4. When a trust caucus strategy is used, mediator impartiality and empathy interact positively to build trust.

Trust caucus strategy and performance indicators

While our focus at this point is on how trust is built rather than how much trust is built, it is still important to link the use of a trust caucus strategy with some performance indicators. What would be the point of using the proposed strategy should it not improve mediation outcomes? On one hand, as shown with numerous classic studies, the most usual performance indicator in workplace mediation is whether or not an agreement was reached (Poitras and LeTareau, 2009). The time needed to reach an agreement is far less studied. Most often, time is used to assess how long a conflict has been going on before mediation starts (Wall and Lynn, 1993). Furthermore, when time is considered in terms of the time needed to reach agreement, it is mostly to compare mediation with other dispute resolution options such as arbitration and litigation (e.g. Emery *et al.*, 1991). However, time is an important performance indicator in a lot of mediation program evaluations. As a result, we have included both agreement and time needed to reach agreement in our research design.

With regard to the likelihood of reaching an agreement, studies on the dynamics of conflict (Rubin *et al.*, 1994; Lewicki and Wiethoff, 2000) clearly indicate that conflict resolution is strongly linked to trust. Yiu *et al.* (2006) found that, based on mediators' perception, the greater the degree of trust, the greater the probability that an agreement

would result from the dialogue. We therefore postulate that the use of a trust caucus strategy will increase the settlement rate:

H5. When a trust caucus strategy is used, the likelihood of reaching an agreement increases.

With regard to the time needed to reach an agreement, trust might reduce the time by facilitating access to key information. Parties' trust in their mediator is important because, without a trust relationship, the mediator may not have access to important information (Butler, 1999; Moore, 2003). As a result, trust is necessary for ensuring that critical information (underlying interest, potential concession, bottom line, etc.) is revealed, in order to move parties toward an agreement. Consequently, it would be logical to assume that a high level of trust would enable a mediator to mediate more efficiently (i.e. gain access to critical information), and thus speed up the mediation process. In fact, a previous research has suggested that trust-building techniques shorten the length of time needed to settle a dispute (Tak Wing and Wai Ying, 2009). Since the trust caucus strategy is likely to increase parties' trust and because trust is linked to the time needed to reach agreement, we postulate that the strategy will decrease the time needed to conclude an agreement:

H6. When a trust caucus strategy is used, the time needed to reach agreement is reduced.

Method
The study was conducted in partnership with the *Commission des normes du travail du Québec* (CNT). This organization's mission is to inform the public about matters surrounding the Labour Standards Act, supervise its application and receive complaints from employees. One of the organization's roles is to achieve agreement between employers and employees with respect to disputes relating to the application of this Act and its regulations. The organization examines grievances dealing with dismissals without cause, prohibited work practices and psychological harassment complaints. The mediation program has been around for 15 years and has an agreement rate varying from 70 to 80 percent, which is within the range of what is normally seen in the field (Boulle and Kelly, 1998).

Quasi-experimental design
A quasi-experimental design is used to assess the impact of the trust caucus strategy. This design was selected instead of an experimental design because we could not divide the mediation cases randomly among the different groups and had to comply with administrative procedures. Mediation cases are thus dealt with in the order in which they are added to the CNT's case load. For each experimental condition, cases are therefore distributed on an administrative basis (i.e. based on the order in which requests for mediation are filed) instead of a pure random distribution. According to the program coordinator, there was no reason to suspect any variation in type of cases across the three time periods. Moreover, there were no parties involves in multiple cases. Thus the assignment of subjects to groups meets the requirement of quasi-experimental design.

Given that the study compares different mediation cases for each experimental group, specifications with separate samples were used. Cases were mediated only once, so the logic is not one of repeated measurement. The unit of analysis is the parties'

perspective. In this design, the mediator is a variable that should be set as a constant. As a result, we made sure that it was the same pool of mediators that conducted the mediation over data collection period. Furthermore, mediators were invited to play a part in the experiment on the basis of two criteria: they had to have at least two years of experience as a mediator, and they had to have been mediating on a full-time basis for at least one year. These criteria were used to select mediators with a relatively stable level of experience. We wanted to avoid the learning curve of very novice mediators (i.e. improvement due to a rapid increase in mastery during the initial mediations in a person's career). As a result, we argue that the mediator variable is held constant in this experimental design. The only thing that will vary is the use or not of the Trust caucus strategy. Figure 1 illustrates the quasi-experimental design used in this study.

The experimental design unfolds in three phases. While it is the same pool of mediators that conduct mediation during the three phases, different cases are reported in each phase (i.e. this is not repeated measurement). The first two groups of mediations are conducted before training. An initial group of mediations is carried out at time O_1 (before training 1). Afterward, a second group of mediations is conducted at time O_2 (before training 2). Mediators are then trained to use the trust caucus strategy at time X (i.e. between time O_2 and O_3). The training lasts a full day: the intervention principles are explained and the strategy is demonstrated before the mediators practice the strategy through simulations. As described in the previous section, the core of the trust caucus strategy consists of an initial joint session emphasizing impartiality; and a follow-up caucus emphasizing empathy. Once training has been completed, a third group of mediations is conducted at time O_3 (after training). It is the only group of mediations conducted after training.

It is important to note that, at the end of the training, all mediators pledge to use the trust caucus strategy with each new case for the next five weeks (i.e. until the end of phase O_3). All mediators reported having respected their commitment to use the strategy and most mediators decided to adopt the strategy beyond the five-week period. Consequently, the trust caucus strategy was not used at time O_1 and O_2, but was used systematically at time O_3. We therefore expect negative interactions between a mediator's impartiality and empathy at time O_1 and O_2, and a positive interaction between these two predictors at time O_3. A shift in the interaction pattern at time O_3 is assumed to result from the use of the trust caucus strategy.

With respect to internal validity, according to Campbell and Stanley (1963), the selected quasi-experimental design controls the following threats: maturation, testing, statistical regression and selection. However, this design is not as efficient when it comes to controlling the effects of history and instrumentation. With respect to history, we made sure that all trainings, conferences and readings that could influence the mediators' performance were controlled. In fact, the training on the trust caucus strategy was the
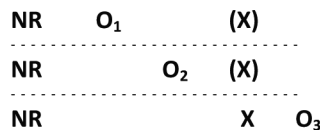

Figure 1.
Pre- and post-test
quasi-experimental design
using separate samples

Notes: NR = non-random assignment; O = Observation;
X = Training; · · · = separate sample
Source: Campbell and Stanley (1963)

only learning activity offered during the period covering the five weeks before the training (O_2 – before training 2) and following the training (O_3 – after training). We thus mitigated the influence of circumstances external to the experiment. To control instrumentation bias, a confirmatory factor analysis of the rating scale was conducted to ensure the reliability of this instrument (Gerbing and Anderson, 1988).

As concerns external validity, note that this quasi-experimental design controls threats to the experiment's external validity (Campbell and Stanley, 1963), which are interaction effects, attrition of participants in a group, and reactive arrangements. The experiment's results may therefore be easier to generalize with respect to other mediation situations. In addition, the realism of the field experiment (in contrast to a laboratory experiment) is a strength of the study for external validity.

Sample
The response rate was 44 percent, which is acceptable for exploratory-type research (Saks *et al.*, 2000). A total of 251 valid questionnaires were kept for statistical analysis ($O_1 = 93$; $O_2 = 74$ and $O_3 = 84$). The sample is composed of employers (51 percent) and employees (49 percent). In terms of gender, 51 percent of the respondents were female and 49 percent were men. Nearly 61 percent have a high school education, 26 percent have an undergraduate university diploma and 13 percent have a graduate level university education. On average, the respondents are 41 years old, with 7 years of seniority in their organization. Most parties work full time (70 percent). The large majority of respondents work in the private sector (72 percent), while the rest work in the public sector. The average salary of respondents is CA$46,031. There was no significant qualitative difference between groups (i.e. data collection periods).

Data collection
Participants are employees and employers involved in a workplace dispute mediated by a professional mediator from the *Commission des normes du travail du Québec* (CNT). Participants were recruited by 24 CNT mediators in eight regions of Quebec. The mediators' role in data collection was to serve as intermediaries between the researchers and parties by making sure the questionnaires were distributed at the end of the mediation according to a predetermined protocol to ensure the scientific validity of the data collection process. Once mediation was completed, mediators were required to inform the parties of the study. In order to avoid bias in selection on the part of the mediators, they had to systematically offer the option of participating to all parties involved in a new mediation case. Parties that agreed to participate responded individually to the questionnaire on site, then sealed it in the prepaid envelope provided to preserve the anonymity of their responses.

To protect the participants' anonymity, the CNT did not allow researchers to code the questionnaire; it was thus not possible to pair parties or link questionnaires with mediators. The data collection period was spread out over three two-month periods: September and October 2006 (O_1 – Before training 1), January and February 2006 (O_2 – Before training 2) and March and April 2006 (O_3 – After training).

Measurement scales
Three scales were developed to evaluate three specific dimensions of parties' trust in their mediators: the mediator's impartiality, the mediator's empathy and the party's general trust in the mediator. Impartiality is defined as giving both parties an equal chance to express themselves without favoring one party over the other. Empathy

means the mediator's general attitude toward the parties, such as warmth and showing consideration. General trust refers to the parties' general feeling regarding the mediator, rather than a specific aspect of trust.

Each scale included two items. For each statement, the parties indicated the degree to which they agreed with the statement, using a Likert-type scale of six points from "disagree strongly" (1) to "agree strongly" (6). The psychometric characteristics of the measurement scales were tested in three steps. First, the results of an exploratory factor analysis show an optimal structure with questions loading significantly only on a single scale (see Table I). Second, the results of a confirmatory factor analysis show a satisfactory construct validity ($p = 0.147$, RMSEA $= 0.033$, NFI $= 0.99$, CFI $= 1.00$, AGFI $= 0.96$). Finally, the degree of homogeneity of the scales was tested using the calculation of inter-item correlations. Results show satisfactory internal consistency for all scales: mediator's impartiality $r = 0.93$, mediator's empathy $r = 0.87$, party's trust in mediator $r = 0.86$.

In addition, parties had to indicate the outcome of the mediation in the questionnaire. Two outcomes were possible: agreement reached or mediation ended without agreement. Consequently, outcome is a binary variable in this research. The parties also had to indicate the approximate duration of their mediation. When the outcome was an agreement, we coded in minutes the time required to reach agreement. When the outcome was not an agreement, the time needed to reach agreement was kept blank. Therefore, only cases that ended in agreement would be used to compute the time needed to reach agreement. Finally, parties had to report their role in mediation (employee or employer). Furthermore, there might be significant difference between groups as to their reaction to the use of the trust caucus strategy by mediators.

Data analysis

The intergroup comparison embedded in a quasi-experimental design enabled us to assess the impact of the trust caucus strategy on key variables. The following linear regression procedures were used for all groups in our sample (before training 1, before training 2 and after training). First, a hierarchical linear regression analysis was used to

	Mediator's impartiality	Component Mediator's empathy	Parties' trust in mediator
(I1) The mediator did not demonstrate any partiality toward either side during the negotiation	0.937		
(I2) The mediator did not show any bias toward either party	0.827		
(E1) I felt that the mediator was compassionate		0.983	
(E2) I felt that the mediator understood me		0.731	
(T1) I felt very confident in the mediator			−0.968
(T2) I had the feeling that I could trust the mediator completely			−0.900

Notes: Extraction method: principal component analysis; Rotation method: Oblimin with Kaiser normalization; Factors loading below 0.30 are not shown because they are not significant

Table I.
Exploratory factor analysis

evaluate the influence of the two mediator characteristics (level of impartiality and empathy) on the parties' level of trust in their mediator while controlling for outcome (i.e. outcome entered first in the equation). We decided to control for outcome because we postulated that the result of a mediation could theoretically impact a party's evaluation of his or her mediator (Ross *et al.*, 2002). Second, we added mediators' impartiality and empathy as predictors of parties' trust in their mediator. Third, we added a regression equation including an interaction effect between the mediator's impartiality and empathy. The overall procedure enabled us to determine whether there is a significant interaction between mediators' impartiality and empathy for all groups. We predicted a negative interaction for the before training 1 and before training 2 groups (O_1 and O_2), but a positive interaction for the after training group (O_3).

Because trust, impartiality and empathy are highly correlated, we centered predictor variables to reduce the impact of multicollinearity. In addition, we splitted the cases according to role (employees and employers) in order to avoid using two evaluations of the same mediator in our sample. As a result, the regression procedure was used for both employers and employees separately. By splitting However, because dividing our sample reduced the statistical power of our analysis, we set the p value to $p < 0.10$ (one tailed p value).

Finally, we conducted two group comparisons between the three conditions (O_1, O_2 and O_3). First, to compare the settlement rate between groups, a Pearson Chi-Square test was used. Second, to compare time needed to reach agreement between the three conditions, an analysis of variance (ANOVA) test was used. The analysis was completed with multiple comparisons based on the Dunnett method, using the before training group 1 as the control. Contrary to regression analysis, we did not run separate group comparisons for employers and employees. This time, the subdivision of our sample reduced too dramatically the statistical power of our analysis. However, we set the p value to $p < 0.01$ to counteract the fact that pairs of disputants from the same mediation rated outcome and time and that this situation might magnify differences in outcome and time needed to reach agreement.

Results
Of the 251 questionnaires, 21 percent of the mediation cases did not result in an agreement, while 79 percent of the mediations resulted in an agreement between the parties. In general, mediators' impartiality ($\mu = 5.26$) and mediators' empathy ($\mu = 5.44$) were rated fairly high. Consequently, it is not surprising that parties' trust in their mediator ($\mu = 5.34$) was also rated high. Furthermore, there is no significant difference of mean for these three variables among the groups (O_1, O_2 and O_3). Finally, note that these three variables were all significantly correlated. However, the likelihood of reaching agreement was not correlated to any variable in the study. Table II highlights intercorrelations of the six variables of the study.

Linear regressions
Linear regression in all conditions (i.e. comparison, before training 2 and after training groups) resulted in significant models (see Tables III and IV). Furthermore, models accounted for about 85 percent of the variance, ranging from 80.6 to 91.6 percent. Consequently, models in each condition have both statistical and practical significance.

With regard to the role of mediators' impartiality and empathy as predictors of parties' trust in their mediator, *H1* and *H2* are both confirmed. For employers, mediators'

impartiality ($\beta_{O1} = 0.59$, $p < 0.001$; $\beta_{O2} = 0.794$, $p < 0.001$; $\beta_{O3} = 0.17$, $p < 0.10$) and empathy ($\beta_{O1} = 0.40$, $p < 0.001$; $\beta_{O2} = 0.17$, $p < 0.10$; $\beta_{O3} = 0.79$, $p < 0.001$) were significant predictors of parties' trust in the mediator in all conditions. For employees, mediators' impartiality ($\beta_{O1} = 0.69$, $p < 0.001$; $\beta_{O2} = 0.61$, $p < 0.001$; $\beta_{O3} = 0.23$, $p < 0.05$) and empathy ($\beta_{O1} = 0.30$, $p < 0.01$; $\beta_{O2} = 0.41$, $p < 0.001$; $\beta_{O3} = 0.67$, $p < 0.001$) were significant predictors of parties' trust in the mediator in all conditions.

Contrary to our prediction, there was no negative interaction between mediators' impartiality and empathy for the before training 1 and the before training 2 groups for employers ($\beta_{O1} = 0.13$, $p > 0.10$; $\beta_{O2} = 0.00$, $p > 0.10$) and employees ($\beta_{O1} = -0.10$, $p > 0.10$; $\beta_{O2} = 0.10$, $p > 0.10$). *H3* is not confirmed. However, there was a positive interaction between impartiality and empathy for the after training groups for both employers ($\beta = 0.19$, $p < 0.05$) and employees ($\beta = 0.49$, $p < .001$). As a result, *H4* is confirmed.

Inter-group comparisons
The cross tabulation of mediation outcomes and time needed to reach agreement with the three experimental groups highlights the significant and practical impact of the trust caucus strategy (see Table V). More specifically, the use of the trust caucus strategy did not significantly improve the agreement rate ($O_1 = 82$ percent, $O_2 = 77$ percent and $O_3 = 76$ percent, $p = 0.63$). As a result, *H5* is not supported. However, there was a significant difference between groups in regard to the time needed to reach agreement ($F = 3.89$, $p = 0.005$). Furthermore, *post hoc* tests showed that there was not a significant difference between the before training 1 and the before training 2 groups ($\Delta\mu = 1.3$, $p = 0.922$) but that there was a significant difference for the after training group ($\Delta\mu = 21.5$, $p < 0.01$) with both before training groups. According to our research design, we therefore conclude that there is a significant difference in time needed to reach agreement when the trust caucus strategy is used. Consequently, *H6* is supported.

Discussion
The findings of the study generate six points of discussion. Firstly, in all groups, mediators' impartiality and empathy are strong factors in determining parties' level of trust in their mediator. Moreover, impartiality and empathy account for on average 80 percent (ranging from 59 percent to 91 percent) of the variance in the parties' trust in the mediator. These results confirm the key role of impartiality and empathy in building the trust relationship between the mediator and parties.

Secondly, we were not able to support the predicted negative interaction (*H3*) between impartiality and empathy in the before training 1 and the before training 2

	Mean	SD	1	2	3	4	5
1. Outcome	0.72	–	–				
2. Parties' trust in mediator	5.34	0.83	0.13	(0.93)			
3. Mediator's impartiality	5.26	0.88	0.12	0.80[*]	(0.87)		
4. Mediator's empathy	5.44	0.91	0.09	0.83[*]	0.73[*]	(0.86)	
5. Time to reach an agreement (min.)	142.8	58.2	0.06	−0.10	−0.12	−0.09	–

Table II.
Descriptive statistics and correlations

Note: [*]Correlation is significant at the 0.01 level (2-tailed); Iter-item correlations are presented in the diagonal

	Before training 1 (O$_1$) $n = 43$			Before training 2 (O$_2$) $n = 39$			After training (O$_3$) $n = 45$		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1: control variables									
Outcome	-0.012	0.000		-0.147	0.021		0.087	0.008	
Step 2: main effects									
Mediator's impartiality	0.593***			0.794***			0.168[a]		
Mediator's empathy	0.396***	0.844***		0.167[a]	0.813***		0.786***	0.783***	
Step 3: interaction									
Impartiality * Empathy	0.133	0.851***	0.007	0.004	0.813***	0.000	0.191*	0.806***	0.022*

Notes: *p < 0.05; **p < 0.01; ***p < 0.001; [a]p < .10

Table III.
Linear regression results – Employers' trust in mediator

	Before training 1 (O$_1$) $n = 50$			Before training 2 (O$_2$) $n = 35$			After training (O$_3$) $n = 39$		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1: control variables									
Outcome	−0.013	0.000		0.020	0.000		0.361*	0.130*	
Step 2: main effects									
Mediator's impartiality	0.685***			0.605***			0.229*		
Mediator's empathy	0.301**	0.912***		0.411***	0.890***		0.669***	0.708***	
Step 3: interaction									
Impartiality * Empathy	−0.098	0.916***	0.004	0.103	0.895***	0.005	0.491***	0.828***	0.119***

Notes: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table IV.
Linear regression results
– employees' trust in
mediator

groups. We propose two hypotheses could explain this result. The first is that our sample was too small to detect significant results. However, this hypothesis is not very compelling because the average interaction for all before training groups is only about 4 percent of the variance. It is very likely that there are no interactions between the two trust factors. The second is that mediators are probably already somewhere aware of the challenge of combining impartiality and empathy and are good at avoiding this pitfall of rapport building. In other words, mediators were able to avoid the interaction between the two factors. The third is that because mediators are very concerned with maintaining an aura of impartiality, perhaps they choose to be more impartial than empathetic. Mediators may have focus most of their effort on being impartial, thus showing very little empathy. By doing so, they avoided the negative interaction between the two factors. This hypothesis would be congruent with our results since the regression weights of impartiality is about twice stronger that the regression weights of empathy in the case mediated before the training. Since in all before training groups, the average regression weight of impartiality (mean $\beta = 0.669$) is about twice as large as the regression weight of empathy (mean $\beta = 0.319$), this hypothesis is probably the most likely explanation.

Thirdly, we did find a positive interaction between impartiality and empathy after the training. It seems that the trust caucus strategy does facilitate building rapport as well as maintaining the perception of the mediator as being fair. However, it is interesting to note that the interaction for employees ($\beta = 0.491$) is about 2.5 times higher than it is for employers ($\beta = 0.191$). Two hypotheses may explain this result. First, it is possible that employees are more negatively sensitive to "seeing" the mediator being empathetic to the employers. As a result, they "benefit" more from the use of the trust caucus strategy which prevents them from seeing mediators demonstrate that they care about the employers' concerns. The second hypothesis is that mediators are more empathetic to employees than they are to employers in caucus and thus create a greater impact. This would be congruent with the findings of Welton *et al.* (1988) who found that mediators usually give greater support in caucus to the side that originally filed the complaint.

Fourthly, we do observe a shift in regression weights between impartiality and empathy after the training. Before the training, the average regression weight of impartiality (mean $\beta = 0.669$) is about twice the regression weight of empathy (mean $b = 0.319$). After the training, the average regression weight of empathy (mean $b = 0.728$) is about 3.6 times higher than the average regression weight of impartiality (mean $\beta = 0.199$). Two hypotheses could explain this result. First, because there are fewer challenges to mediators' impartiality with the use of caucus, mediators focus more on empathy as a key to rapport building. Second, because mediators' impartiality is not the object of parties' scrutiny with the use of trust strategy, parties put more weight on the mediators' empathy in evaluating their trustworthiness. This hypothesis

Variable	Comparison group	Before training	After training	Level of significance
Agreement rate	81.7%	77.0%	76.2%	$p = 0.628$ ($X^2 = 0.932$)
Time to reach agreement (min.)	149.4 (59.3)	150.7 (54.2)	128.6 (54.2)	$p = 0.022^*$ (F = 3.89)

Notes: $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; Standard deviations are in parenthesis

Table V.
Intergroup comparisons

resonates with Mayer *et al.*'s (1995) proposition that empathy (benevolence) is eventually more important than impartiality (integrity) over time. Therefore, the trust caucus strategy would accelerate the process of putting more weight on empathy as a key component of trust.

Fifthly, we did not observe a significant increase in the likelihood of reaching an agreement when the trust caucus strategy was used. This result is congruent with the Weldon *et al.* (1992) findings that the use of caucus is not associated with an increase in the likelihood of reaching an agreement. The authors argue that, while the use of caucus has advantages, it also has disadvantages. They found that the use of caucus increases indirect hostility among parties (e.g. putting down the other, providing exaggerated negative information) and thus makes reaching an agreement harder. Thus, using a caucus strategy would not improve the likelihood of reaching an agreement because its disadvantages offset its advantages. A second hypothesis is that, while the trust caucus strategy does improve trust, it does not increase the likelihood of reaching an agreement because there is no linear relationship between the two variables. A recent study has found that there is a threshold relationship between trust and the likelihood of reaching an agreement (Stimec and Poitras, 2009). The study's result showed that, after a minimum level of trust is reached, any increase in trust does not translate into a higher settlement rate. A final explanation is that there might simply not be enough variation in trust levels to produce variation in agreement rates. Overall, it seems that the trust caucus strategy does not increase trust, but, as we postulated in the previous point, changes how trust is built.

Sixthly, we did observe a significant decrease in the time needed to reach agreement. It is surprising that the strategy reduces mediation time when it does not impact the likelihood of success. One hypothesis is that the trust caucus strategy does not improve the level of trust, but speeds up the trust building process. By adding a synergetic effect between impartiality and empathy, parties develop trust in their mediation more quickly. A second hypothesis is that by focusing more on empathy in caucus, mediators get deeper access to information. Welton *et al.* (1988) have found that parties provided more information in caucus than in joint session. It would be this deeper access to information that speeds up the mediation process. Because the shifts in regression weight between impartiality and empathy are more dramatic than the interactions between the two variables, the last hypothesis is the most likely.

Practical and theoretic implications

Our findings confirm what most mediators know intuitively, which is that using caucus to build trust is an effective strategy. When mediators use the trust caucus strategy, impartiality and empathy work better together and parties put more weight on empathy than on impartiality. While the use of the trust caucus does not increase the likelihood of reaching agreement, it does significantly decrease the time needed to conclude an agreement. One potential explanation is that by focusing on empathy, mediators gain access to key information more quickly, thus speeding up the mediation process.

The findings have two theoretic implications. First, while several authors have already postulated the key role of empathy in rapport building, this study goes one step further. The results suggest that empathy might ultimately be more important than impartiality in establishing mediators' trustworthiness. Second, the results show the importance of adding time needed to reach agreement as a performance indicator. If time had not been included in this study, we would have assumed that the trust caucus

strategy did not yield any improvement in performance. While outcome (agreement/no agreement) is the most classic performance indicator in mediation, it might be wise to include time needed to conclude an agreement as well.

Conclusion

Mediator impartiality and empathy are two classic factors in the trust relationship between mediator and parties. Unsurprisingly, this study confirmed the key role of these two factors. Although the theoretical antagonistic relationship between these two factors was not supported by our findings, we did find a synergetic relationship between the factors when impartiality and empathy are separated using a trust caucus strategy. Furthermore, the main practical advantage of this strategy was to reduce the time needed to reach agreement. However, the trust caucus strategy did not improve parties' trust in the mediator, nor did it improve the settlement rate. These findings have prompted us to hypothesize that the trust caucus strategy does not improve the level of trust, but rather speeds up the process of building a rapport between the mediator and parties.

Our study has three main limitations. First, all our data come from workplace mediation and from the same organization. Although we can reasonably postulate that the results can be generalized to other mediation settings, this remains to be proven. Second, it is important to note that the causal relationship between the trust caucus strategy and the time needed to conclude an agreement is only inferred. The use of caucus may possibly speed up the mediation process independent of its impact on the trust relationship. For example, using caucus might cut short unproductive conversation between parties at the beginning of mediation. Nevertheless, there was an interaction between impartiality and empathy when the trust caucus strategy was used and there was also a important shift in regression weight between the two variables. Third, trust, rapport, and impartiality are all subjective measures obtained from the same respondent. Hence a strong degree of common factor bias may have been introduced into the statistical results.

In future studies, it might be interesting to identify which mediating variable explains how parties' trust is linked to time required to conclude an agreement. One potential variable would be the mediator's access to critical information regarding the parties. Greater trust in a mediator means he or she has easier access to information to help guide parties toward a potential agreement. The faster a mediator gains the trust of parties, the faster he or she has access to information and the faster the mediation can be concluded.

References

Albin, C. (1993), "The role of fairness in negotiation", *Negotiation Journal*, Vol. 9 No. 3, pp. 223-44.

Augsburger, D.W. (1992), *Conflict Mediation Across Culture: Pathways and Patterns*, Westminster/John Knox Press, Louisville, KY.

Boulle, L. and Kelly, K.J. (1998), *Mediation: Principles, Process, Practice*, Butterworths Canada, Toronto.

Butler, J.K. (1991), "Toward understanding and measuring conditions of trust: evolution of a conditions of trust inventory", *Journal of Management*, Vol. 17, pp. 643-63.

Butler, J.K. (1999), "Trust expectations, information sharing, climate of trust and negotiation effectiveness and efficiency", *Group & Organization Management*, Vol. 24 No. 2, pp. 217-38.

Campbell, D.T. and Stanley, J.C. (1963), *Experimental and Quasi-Experimental Designs for Research*, Houghton Miffin Company, Boston, MA.

Desivilya, H. and Rotem, M.R. (2008), "Do mediators walk the talk?: a study of Israeli mediators", *Dispute Resolution Journal*, Vol. 63 No. 1, pp. 63-70.

Deutsch, M. (1958), "Trust and suspicion", *Journal of Conflict Resolution*, Vol. 2 No. 4, pp. 265-79.

Emery, R.E., Matthews, S.G. and Wyer, M.M. (1991), "Child custody mediation and litigation: further evidence on the differing views of mothers and fathers", *Journal of Consulting and Clinical Psychology*, Vol. 59 No. 3, pp. 410-8.

Gerbing, D.W. and Anderson, J.C. (1988), "An updated paradigm for scale development incorporating unidimensionality and its assessment", *Journal of Marketing Research*, Vol. 25 No. 2, pp. 186-92.

Goldberg, S.B. (2005), "The secrets of successful mediators", *Negotiation Journal*, Vol. 21 No. 3, pp. 365-76.

Honeyman, C. (1988), "Five elements of mediation", *Negotiation Journal*, Vol. 4 No. 2, pp. 149-52.

Kelman, H.C. and Cohen, S.P. (1979), "Reduction of international conflict: an interactional approach", in Austin, W.G. and Worchel, S. (Eds), *The Social Psychology of Intergroup Relation*, Brooks/Cole, Monterey, CA, pp. 288-303.

Kolb, D. (1985), "To be a mediator: expressive tactics in mediation", *Journal of Social Issues*, Vol. 41 No. 2, pp. 11-26.

Kressel, K. (1972), *Labor Mediation: An Exploratory Survey*, Association of Labor Mediation Agencies, Albany, NY.

Lewicki, R.J. and Wiethoff, C. (2000), "Trust, trust development, and trust repair", in Deutsch, M. and Coleman, P.C. (Eds), *The Handbook of Conflict Resolution: Theory and Practice*, Jossey-Bass Publishers, San Francisco, CA, pp. 86-107.

McCarthy, W. (1985), "The role of power and principle in getting to yes", *Negotiation Journal*, Vol. 2 No. 1, pp. 59-66.

McKnight, H.D. and Chervany, N.L. (2001), "Trust and distrust definitions: one bite at a time", in Falcone, R., Singh, M.P. and Tan, Y.H. (Eds), *Trust in Cyber-Societies*, Springer, New York, NY, pp. 27-54.

Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995), "An integrative model of organizational trust", *Academy of Management Review*, Vol. 20 No. 3, pp. 709-34.

Mishra, A.K. (1996), "Organizational response to crisis: the centrality of trust", in Kramer, R.M. and Tylers, T. (Eds), *Trust in Organizations*, Sage, Newbury Park, CA, pp. 261-87.

Moore, C.W. (2003), *The Mediation Process: Practical Strategies for Resolving Conflict*, 3rd ed., Jossey Bass, San Francisco, CA.

Poitras, J. and LeTareau, A. (2009), "Quantifying the quality of mediation agreements", *Negotiation and Conflict Management Research*, Vol. 2 No. 4, pp. 363-80.

Pruitt, D.G. and Lewis, S.A. (1983), "Achieving integrative agreements", in Bazerman, M. and Lewicki, R.J. (Eds), *Negotiation in Organizations*, Sage, Thousand Oaks, CA.

Ross, W.H., Brantmeier, C. and Ciriacks, T. (2002), "The impact of hybrid dispute-resolution procedures on constituent fairness judgments", *Journal of Applied Social Psychology*, Vol. 32 No. 6, pp. 1151-88.

Rubin, J.Z., Pruitt, D.G. and Kim, S.H. (1994), *Social Conflict: Escalation, Stalemate, and Settlement*, 2nd ed., McGraw-Hill, New York, NY.

Saks, A.M., Schmitt, N.W. and Kilmoski, R.J. (2000), *Research, Measurement, and Evaluation of Human Resources*, Scarborough, Nelson.

Sherman, M. (2003), "Mediation hype and hyperbole: how much should we believe?", *Dispute Resolution Journal*, Vol. 58 No. 3, pp. 43-51.

Stimec, A. and Poitras, J. (2009), "Building trust with parties: are mediators overdoing it?", *Conflict Resolution Quarterly*, Vol. 26 No. 3, pp. 317-31.

Strickland, L.H. (1958), "Surveillance and trust", *Journal of Personality*, Vol. 26, pp. 200-15.

Tak Wing, Y. and Wai Ying, L. (2009), "Efficacy of trust-building tactics in construction mediation", *Journal of Construction Engineering and Management*, Vol. 135 No. 8, pp. 683-9.

Vallone, R.P., Ross, L. and Lepper, M.R. (1985), "The hostile media phenomenon: biased perception and perception of media bias in coverage of the Beirut massacre", *Journal of Personality and Social Psychology*, Vol. 49, pp. 577-85.

Wall, J. and Lynn, A. (1993), "Mediation: a current review", *Journal of Conflict Resolution*, Vol. 37 No. 1, pp. 160-94.

Wehr, P. and Lederach, J.P. (1991), "Mediating conflict in central America", *Journal of Peace Research*, Vol. 28 No. 1, pp. 85-98.

Welton, G.L. and Pruitt, D.G. (1987), "The mediation process: the effect of mediator bias and disputant power", *Personality and Social Psychology Bulletin*, Vol. 13 No. 1, pp. 123-33.

Welton, G.L., Pruitt, D.G. and McGillicuddy, N.B. (1988), "The role of caucusing in community mediation", *Journal of Conflict Resolution*, Vol. 32 No. 1, pp. 181-202.

Yiu, T.W., Cheung, S.O. and Mok, F.M. (2006), "Logistic likelihood analysis of mediation outcome", *Journal of Construction Engineering and Management*, p. 1026.

Zaheer, A., McEvily, B. and Perrone, V. (1998), "Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance", *Organization Science*, Vol. 9 No. 2, pp. 141-59.

Zubeck, J., Pruitt, D., Peirce, R., McGillicuddy, N. and Syna, H. (1992), "Disputant and mediator behaviors affecting short-term success", *Journal of Conflict Resolution*, Vol. 36 No. 3, pp. 546-72.

About the author
Jean Poitras is an Associate Professor of Conflict Management at HEC Montréal. He developed comprehensive knowledge of conflict management through his experience as a consultant and the results of his research projects. For more than 15 years now, he has been helping professionals develop their negotiation and mediation skills. His research program focuses on the intervention strategies used in mediation and negotiation, as well as on the development of diagnostic tools to measure the levels of conflict in organizations. Jean Poitras can be contacted at: jean.poitras@hec.ca

CHICAGO

The Law of Group Polarization

Cass R. Sunstein

**THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO**

The Law of Group Polarization

Cass R. Sunstein[*]

Abstract

In a striking empirical regularity, deliberation tends to move groups, and the individuals who compose them, toward a more extreme point in the direction indicated by their own predeliberation judgments. For example, people who are opposed to the minimum wage are likely, after talking to each other, to be still more opposed; people who tend to support gun control are likely, after discussion, to support gun control with considerable enthusiasm; people who believe that global warming is a serious problem are likely, after discussion, to insist on severe measures to prevent global warming. This general phenomenon -- group polarization -- has many implications for economic, political, and legal institutions. It helps to explain extremism, "radicalization," cultural shifts, and the behavior of political parties and religious organizations; it is closely connected to current concerns about the consequences of the Internet; it also helps account for feuds, ethnic antagonism, and tribalism. Group polarization bears on the conduct of government institutions, including juries, legislatures, courts, and regulatory commissions. There are interesting relationships between group polarization and social cascades, both informational and reputational. Normative implications are discussed, with special attention to political and legal institutions.

"The differences of opinion, and the jarrings of parties in [the legislative] department of the government . . . often promote deliberation and circumspection; and serve to check the excesses of the majority."

Alexander Hamilton[1]

"In everyday life the exchange of opinion with others checks our partiality and widens our perspective; we are made to see things form the standpoint of others and the limits of our vision are brought home to us. . . . The benefits from discussion lie in the fact that even representative legislators are limited in knowledge and the ability to reason. No one of them knows everything the others know, or can make all the same inferences that they can draw in concert. Discussion is a way of combining information and enlarging the range of arguments."

John Rawls[2]

---

[1] The Federalist No. 70, at 426-37 (Alexander Hamilton) (Clnton Rossiter ed. 1961).
[2] John Rawls, A Theory of Justice 358-59 (1971).

"Each person can share what he or she knows with the others, making the whole at least equal to the sum of the parts. Unfortunately, this is often not what happens . . . . As polarization gets underway, the group members become more reluctant to bring up items of information they have about the subject that might contradict the emerging group consensus. The result is a biased discussion in which the group has no opportunity to consider all the facts, because the members are not bringing them up. . . . Each item they contributed would thus reinforce the march toward group consensus rather than add complications and fuel debate."

Patricia Wallace[3]

Consider the following events:

- Affirmative action is under attack in the state of Texas. A number of professors at a particular branch of the University of Texas, inclined to be supportive of affirmative action, meet to exchange views and to plan further action, if necessary. What are these professors likely to think, and to do, after they talk?
- After a nationally publicized shooting at a high school, a group of people in the community, most of them tentatively in favor of greater gun control, come together to discuss the possibility of imposing new gun control measures. What, if anything, will happen to individual views as a result of this discussion?
- A local group of citizens, all of them Republicans, meet in 1998 to discuss whether President Clinton should be impeached. Before discussion begins, a strong majority is leaning in favor of impeachment, but they are not firmly committed to this view. A minority is entirely undecided. If a group resolution is required, what is it likely to look like?
- A jury is deciding on an appropriate punitive damage award in a case of recklessly negligent behavior by a large company; the behavior resulted in a serious injury to a small child. Before deliberating as a group, the jurors have chosen appropriate awards, leading to an average of $1.5 million and a median of $1 million. As a statistical generalization, how will the jury's ultimate award tend to compare to these figures?
- A group of women are concerned about what they consider to be a mounting "tyranny of feminism." They believe that women should be able to make their own choices, but they also think that men and women are fundamentally different, and that their differences legitimately lead to different social roles. The group decides to meet every two weeks to focus on common concerns. After a year, is it possible to say what its members are likely to think?
- There is an Internet discussion group, consisting of people concerned about the behavior of certain activities by Americans apparently associated with

---

[3] Patricia Wallace, The Psychology of the Internet 81-82 (1999).

China. Over half of the participants are fearful that China might be engaged in spying and that under President Clinton, the Department of Justice has turned a blind eye, in part because of campaign contributions from Americans whose loyalties are suspect. In what directions are these Internet discussions likely to lead?

Every society contains innumerable deliberating groups. Faculties, juries, legislative bodies, political organizations, regulatory commissions, multimember courts, faculties, student organizations, religious sects, Internet discussion groups, and others engage in deliberation. A pervasive question has to do with the likely consequences of the deliberative process. It is a simple social fact that sometimes people enter discussions with one view and leave with another, even on political and moral questions.[4] Emphasizing this fact, many recent observers have embraced the traditional American aspiration to "deliberative democracy," an ideal that is designed to combine popular responsiveness with a high degree of reflection and exchange among people with competing views.[5] But for the most part, the resulting literature has not been empirically informed.[6] It has not much dealt with the real-world consequences of deliberation, and with whether any generalizations hold in actual deliberative settings.

The standard view of deliberation is that of Hamilton and Rawls, as stated above. Group discussion is likely to lead to better outcomes, if only because competing views are stated and exchanged. Aristotle spoke in similar terms, suggesting that when diverse groups "all come together . . . they may surpass – collectively and as a body, although not individually – the quality of the few best. . . . When there are many who contribute to the process of deliberation, each can bring his share of goodness and moral prudence; . . . some appreciate one part, some another, and all together appreciate all."[7] An important question is whether this view is naïve or excessively optimistic. Perhaps economic, psychological, and social mechanisms lead deliberating groups in unexpected and undesirable directions. If so, it would be necessary to rethink current enthusiasm for deliberation as a social phenomenon, and also to reassess and perhaps to restructure institutions that are designed as deliberating bodies.

My principal purpose in this Article is to investigate a striking but thus far almost entirely neglected[8] empirical regularity – that of **group polarization** -- and to relate this phenomenon to a number of issues in law and political theory**.** In brief, group polarization arises when **members of a deliberating group move toward a more extreme point in whatever direction is**

---

[4] Sometimes it may seem that moral and political arguments are unlikely to have an effect; the evidence discussed here shows that on this proposition is quite wrong as an empirical matter.

[5] See Amy Gutmann and Dennis Thompson, Democracy and Disagreement (1997); Deliberative Democracy (Jon Elster ed. 1998); Jurgen Habermas, Between Law and Norms (1997).

[6] Exceptions include James Fearon, Deliberation As Discussion, in Deliberative Democracy, supra, at 44; Susan Stokes, Pathologies of Deliberation, in id. at 123; Lynn Sanders, Against Deliberation, Political Theory. Of special interest is James Fishkin's continuing experiments with the "deliberative opinion poll," in which groups of diverse people are asked to deliberate on public issues. See James Fishkin, The Voice of the People (1998); The Poll With A Human Face (Maxwell McCombs and Amy Reynolds eds. 1999). Fishkin's groups do not polarize, at least not systematically; this result is undoubtedly a product of the distinctive setting, in which materials are presented on each issue, with corresponding claims of fact and value. In the experiments discussed here, the relevant arguments are introduced by the participants, not by any third party. See below for discussion of Fishkin.

[7] Aristotle, Politics 123 (E. Barker trans. 1972).

[8] I have been unable to find any sustained discussions in the relevant literature in law or political theory.

indicated by the members' predeliberation tendency. "[L]ike polarized molecules, group members become even more aligned in the direction they were already tending."[9] Group polarization is the conventional consequence of group deliberation. Thus, for example, the first deliberating group is likely to become more firmly committed to affirmative action; the second group will probably end up favoring gun control quite enthusiastically; any group resolution from the third group will tend to favor impeachment; the punitive damages jury will likely come up with an award higher than the median and perhaps higher than the mean as well; the group of women concerned about feminism is likely to become very conservative indeed on gender issues; the Internet group is likely to fear something like a conspiracy to cover up the relevant activities.[10]

Two principal mechanisms underlie group polarization. The first points to social influences on behavior; the second emphasizes limited "argument pools," and the directions in which those limited pools lead group members. An understanding of these mechanisms provides many insights into legal and political issues; it illuminates a great deal, for example, about likely processes within multimember courts, juries, political parties, and legislatures – not to mention insulated ethnic groups, extremist organizations, student associations, faculties, workplaces, and families. At the same time, these mechanisms give little reason for confidence that deliberation is making things better than worse; in fact they raise some serious questions about deliberation from the normative point of view.[11] If deliberation simply pushes a group toward a more extreme point in the direction of its original tendency, do we have any systematic reason to think that discussion is producing improvements?

As we will see, one of the principal lessons of the group polarization phenomenon is to cast new light on an old point, to the effect that social homogeneity can be quite damaging to good deliberation.[12] When people are hearing echoes of their own voices, the consequence may be far more than support and reinforcement. Another lesson is that particular forms of homogeneity can be breeding grounds for unjustified extremism, even fanaticism. To work well, deliberating groups should be appropriately heterogeneous and should contain a plurality of articulate people with reasonable views – an observation with implications for the design of regulatory commissions, legislative committees, White House working groups, and even multimember courts.[13] But there is a conceptual problem here: It is difficult to specify appropriate heterogeneity, and the appropriate plurality of views, without making some antecedent judgments about the substantive question at issue. I offer some comments about how to resolve that problem.

This Article is organized as follows. Part II offers some brief notations on the general question of social influences on individual judgments, with particular reference to the

---

[9] See John Turner et al., Rediscovering the Social Group 142 (1987).

[10] Compare R. Hightower and L. Sayeed, The Impact of Computer-Mediated Communication Systems on Biased Group Discussion, 11 Computers in Human Behavior 33 (1995).

[11] I am speaking here of real-world deliberation, not of deliberation accompanied by preconditions of the sort that have been influenced by those thinking of it in ideal terms. See Jurgen Habermas, supra. A particular point to emphasize here is the need for full information. See id.

[12] A classic discussion is John Stuart Mill, On Liberty (1859).

[13] Compare Irving Janis, Groupthink (1972) (coming to the same general conclusion, but without discussing polarization).

phenomenon of social cascades. A central point here is that any particular's persons deeds and statements create an <u>informational externality</u>.[14] When a number of people have acted or spoken, observers who lack much private information are highly likely to follow their lead. Part III offers a basic account of group polarization, with particular reference to some new data in the legal context. Part IV discusses the mechanisms that account for group polarization. Part V traces the implications for a number of issues, involving feuds, ethnic strife, juries, commissions, multimember courts, legislatures, and deliberation via the Internet. Part VI shows in what sense group polarization raises doubts about the idea that deliberation is a social good; it traces the implications of the phenomenon for proper structuring of deliberative institutions. Part VII is a brief conclusion.

II.   Social Influences and Cascades

A. In General

A great deal of attention has recently been devoted to the topic of social influences on individual behavior.[15] Because many of these influences are at least roughly analogous to what happens in group polarization, and because they have some bearing on deliberation as well, it will be worthwhile to offer some brief notations here.

The simplest point is that people frequently do what they do because of what they think (relevant) others do. Thus, for example, teenage girls who see that other teenagers are having babies are more likely to become pregnant themselves[16]; littering and nonlittering behavior appears to be contagious[17]; the same is true of violent crime[18]; those who know other people who are on welfare are more likely to go on welfare themselves[19]; the behavior of proximate others affects the decision whether to recycle[20]; a good way to increase the incidence of tax compliance is to inform people of high levels of voluntary tax compliance[21]; and students are less likely to engage in binge drinking if they think that most of their fellow students do not engage in binge drinking, so much so that disclosure of this fact is one of the few successful methods of reducing binge drinking on college campuses.[22]

---

[14] See Andrew Caplin & John Leahy, Miracle on Sixth Avenue: Information Externalities and Search, 108 Econ. J. 60 (1998).

[15] See, e.g., Dan Kahan, Social Influence, Social Meaning, and Deterrence, Va L Rev. (1998). For extended overviews., see Eliott Aronson, The Social Animal (7th ed. 1995); Lee Ross and Richard Nisbett, The Person and the Situation (1991); group polarization is a surprising omission from both of these lengthy and highly illuminating treatments.

[16] *See, e.g.*, George A. Akerlof, Janet L. Yellen & Michael L. Katz, An Analysis of Out-of-Wedlock Childbearing in the United States, 111 Q.J. Econ. 277 (1996).

[17] See Robert Cialdini et al., A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places, 58 J Pers. And Soc. Psych 1015 (1990).

[18] See Washington Post (December 1999).

[19] See Marianne Bertrand, Erzo F.P. Luttmer & Sendhil Millainathan, Network Effects and Welfare Cultures (unpublished manuscript, Apr. 9, 1998).

[20] See Ardith Spence, Wants for Waste (unpublished Ph.D. dissertation, University of Chicago, 1999).

[21] See Stephen Coleman, The Minnesota Income Tax Compliance Experiment State Tax Results (Minnesota Department of Revenue, April 1996).

[22] See H. Wesley Perkins, College Student Misperceptions of Alcohol and Other Drug Norms Among Peers, in Designing Alcohol and Other Drug Prevention Programs in Higher Education 177-206 (US Dept.of Educ. ed. 1997); Timur Kuran and Cass R. Sunstein, Availability Cascades and Risk Regulation, 51 Stan L Rev 683, 767 (1999).

Social influences affect behavior via two different mechanisms.[23] The first is informational. As noted, what other people do, or say, carries an informational externality; if many other people go to a certain movie, or refuse to use drugs, or carry guns, observers are given a signal about what it makes sense to do. The second mechanism is reputational. Even if people do not believe that what other people do provides information about what should be done, they may think that the actions of others provide information about what other people <u>think</u> should be done. Thus each person's expressive actions come with a <u>reputational externality</u>. People care about their reputations, and hence they may do what they think other people think they should do, whether or not they believe that they should do it. Reputational considerations may, for example, lead people to obey or not to obey the law, smoke cigarettes, buy certain cars,[24] drive while drunk, help others, or talk about political issues in a certain way. They exert a ubiquitous influence on behavior.[25]

B. Some Classic Experiments

In the most vivid experiments involving group influences, conducted by Solomon Asch, individuals were willing to abandon the direct evidence of their own senses.[26] In the relevant experiments, a certain line was placed on a large white card. The task of the subjects was to "match" that line by choosing, as identical to it in length, one of three other lines, placed on a separate large white card. One of the lines on the second white card was in fact identical in length to the line to be matched to it; the other two were substantially different, with the differential varying from an inch and three quarters to three quarters of an inch. The subject in the experiments was one of eight people asked to engage in the matching. But unbeknownst to the subject, the other people apparently being tested were actually there as part of the experiments.

Asch's experiments unfolded in the following way. In the first two rounds, everyone agreed about the right answer; this seemed to be an extremely dull experiment. But the third round introduced "an unexpected disturbance,"[27] Other group members made what was obviously, to the subject and to any reasonable person, a clear error; they matched the line at issue to one that was obviously longer or shorter. In these circumstances the subject had the choice of maintaining his independent judgment or instead yielding to the crowd. A large number of people ended up yielding. In ordinary circumstances subjects erred less than 1 percent of the time; but in rounds in which group pressure supported the incorrect answer, subjects erred 36.8% of the time.[28] Indeed, in a series of twelve questions, no less than 70% of subjects went along with the group, and defied the evidence of their own senses, at least once.

---

[23] See, e.g., Elliott Aronson, supra note, at 22; Lee and Ross, supra note, at 44-45.

[24] See Robert Frank, Luxury Fever (1999).

[25] Timur Kuran, Public Truth, Private Lies (1998), emphasizes this point.

[26] See the overview in Solomon Asch, Opinions and Social Pressure, in Readings About the Social Animal 13 (Elliott Aronson ed. 1995).

[27] Id. at 15.

[28] Id. at 16.

Several refinements are important here. Susceptibility to group influence was hardly uniform; some people agreed with the group almost all of the time, whereas others were entirely independent in their judgments. Significantly, the existence of at least one compatriot, or voice of sanity, mattered a great deal. When just one other person made an accurate match, errors were reduced by three-quarters, even if there was a strong majority the other way.[29] By contrast, varying the size of the majority mattered only up to a number of three, and increases from that point had little effect. Thus opposition from one person did not increase subjects' errors at all; opposition from two people increased error to 13.6%; and opposition from three people increased error to 31.8%, not substantially different from the level that emerged from further increases in group size.

Both informational and reputational considerations appear to have led people toward these errors. Several people said, in private interviews, that their own opinions must have been wrong. On the other hand, experimenters find greatly reduced error, in the same basic circumstances as Asch's experiments, when the subject is asked to give a purely private answer.[30]

Asch concluded that his results raised serious questions about the possibility that "the social process is polluted" by the "dominance of conformity."[31] He added, "That we have found the tendency to conformity in our society so strong that reasonably intelligent and well-meaning young people are willing to call white black is a matter of concern."[32] Notably, however, Asch's experiments did not involve deliberation, for people were not exchanging reasons; indeed, we might expect that reason-giving would have severely weakened his results. What reasons could have been given for incorrect matches? But the existence of substantial numbers of mistakes, as a result of mere exposure to the incorrect conclusions of others, raises questions about whether and when deliberation will lead people in the right directions.

C. Cascades

Some of the most interesting recent work on social influence involves the possibility of informational and reputational "cascades"[33]; this work has obvious relevance to law and politics.[34] Indeed, it is possible to interpret Asch's work as having demonstrated considerable individual susceptibility to cascade effects. What is striking about such effects is that their ripple-like nature, or the quality of contagion.   Group polarization is sometimes, but not always, a product of cascade effects; it will be useful to understand the former against the background of the latter.

The question explored in the cascades literature is why individuals and social groups sometimes move quite rapidly in some direction or another. A starting point is that when individuals lack a great deal of private information (and sometimes even when they have

---

[29] Id. at 18.

[30] See Aronson, supra note, at 23-24.

[31] Id. at 21.

[32] Id.

[33] See Sushil Biikhchandani et al., Learning from the Behavior of Others, J. Econ. Persp., Summer 1998, at 151

[34] See id.; Timur Kuran and Cass R. Sunstein, Availability Cascades and Risk Regulation, 51 Stan L Rev (1999).

such information), they tend to rely on information provided by the statements or actions of others. If A is unaware whether abandoned toxic waste dumps are in fact hazardous, he may be moved in the direction of fear if B seems to think that fear is justified. If A and B believe that fear is justified, C may end up thinking so too, at least if she lacks independent information to the contrary. If A, B, and C believe that abandoned hazardous waste dumps are hazardous, D will have to have a good deal of confidence to reject their shared conclusion. The result of this process can be to produce cascade effects, as large groups of people end up believing something – even if that something is false – simply because other people seem to believe it too. There is a great deal of experimental evidence of informational cascades, which are easy to induce in the laboratory[35]; real world phenomena also seem to have a great deal to do with cascade effects.[36] Notice here that when a cascade is occurring, large numbers of persons end up with a shared view, not simply because of social influence, but via a particular process, in which a rivulet ends up as a flood; this is what makes cascades distinctive.

Though the cascades phenomenon has largely been discussed in connection with factual judgments, the same processes should be at work for political, legal, and moral questions; we can easily imagine political, legal, and moral cascades. Suppose, for example, that A believes that affirmative action is wrong, that B is otherwise in equipoise but shifts upon hearing what A believes, that C is unwilling to persist in his modest approval of affirmative action when A and B disagree; it would be a very confident D who would reject the moral judgments of three (apparently) firmly committed others. Sometimes people are not entirely sure whether capital punishment should be imposed, whether the Constitution protects the right to have an abortion, whether it is wrong to litter or to smoke. Many people, lacking firm convictions of their own, may end up believing what (relevant) others seem to believe. Recent changes in social attitudes toward smoking, recycling, and sexual harassment have a great to do with these effects.[37] The same process may work for the choice of political candidates, as a fad develops in favor of one or another – a cascade "up" or "down," with sensational or ruinous consequences. We can easily imagine cascade effects in the direction of certain judgments about the appropriate course of constitutional law; indeed such effects seem to have been at work in the legal culture in the 1960s (with mounting enthusiasm for the Warren Court) and the 1980s (with mounting skepticism about that Court). It is even possible to imagine cascade effects with respect to questions of constitutional method (eg, textualism, originalism).

Thus far the discussion has involved purely informational pressures and informational cascades, where people care about what other people think because they do not know what to think, and they rely on the opinions of others, to show what it is right to think. But there can be reputational pressures and reputational cascades as well.[38] Here the basic idea is that people care about their reputations, and they speak out, or remain silent, or even engage in certain expressive activity, partly in order to preserve those reputations, even at the price of failing to say what they really think. Suppose, for example, that A believes that hazardous waste dumps pose a serious

---

[35] See Lisa Anderson and Charles Holt, Information Cascades in the Laboratory, 87 Am Econ Rev 847 (1997).

[36] See Bikhchandani et al., A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades, 100 J Polit Econ. 992 (1992); Kuran and Sunstein, supra note.

[37] See Spence, supra note; Cass R. Sunstein, Social Norms and Social Roles, in Cass R. Sunstein, Free Markets and Social Justice ch. 2 (1997).

[38] See Timur Kuran, Public Lies, Private Truths (1996).

environmental problem; suppose too that B is skeptical. B may keep quiet, or (like some of Asch's subjects) even agree with A, simply in order to preserve A's good opinion. C may see that A believes that hazardous waste dumps pose a serious problem, and that B seems to agree with A; C may therefore voice agreement even though privately she is skeptical or ambivalent. It is easy to see how this kind of thing might happen with in political life with, for example, politicians expressing their commitment to capital punishment (even if they are privately skeptical) or their belief in God (even if they are agnostic on the question). Here too the consequence can be cascade effects – large social movements in one direction or another -- when a number of people appear to support a certain course of action simply because others (appear to) do so. What is true for factual beliefs can be true as well for moral, legal, and political judgments. People might say, for example, that affirmative action violates the Constitution simply because of perceived reputational sanctions from saying the opposite; they might support or oppose the death penalty largely in order to avoid the forms of social opprobrium that might come, in the relevant community, from taking the opposing view.

Are social cascades good or bad? No general answer would make sense. Sometimes cascades are quite fragile, precisely because people's commitments are based on little private information; sometimes cascades are rooted in (and greatly fuel) blunders. Sometimes cascade effects will eliminate public torpor, by generating concern about serious problems; but sometimes cascade effects will make people far more worried than they be, or otherwise produce large-scale distortions in private judgments, public policy, and law. The antislavery movement had distinctive cascade-like features, as did the anti-apartheid movement in South Africa; so too with Mao's Cultural Revolution and the rise of Nazism in Germany.[39] The serious risk with social cascades, both informational and reputational, is that they can lead to widespead errors, factual or otherwise. Cascades need not involve deliberation; but related problems infect processes of group deliberation, as we will now see.

III. How and Why Groups Polarize

A. The Basic Phenomenon

Group polarization is among the most robust patterns found in deliberating bodies, and it has been found in many diverse tasks. Polarization is said "to occur when an initial tendency of individual group members toward a given direction is enhanced [by] group discussion."[40]   The result is that groups often make more extreme decisions than would the typical or average individual in the group (where "extreme" is defined internally, by reference to the group's initial dispositions). There is a clear relationship between group polarization and cascade effects; as we will see, the former, like the latter, seems to have a great deal to do with both informational and reputational influences. A key difference is that cascade effects lead people to fall in line with an existing tendency, whereas polarization leads them to a more extreme point in the same direction.

---

[39] See The Social Life of Nazi Germany (1999).
[40] See Isenberg, supra note, at 1141.

Notice that group polarization refers not to variance among groups of any kind, but to what happens <u>within</u> a group discussing a case or problem.[41] Consider some examples of the basic phenomenon, which has been found in an array of nations.[42] (a) A group of moderately profeminist women will become more strongly profeminist after discussion.[43] (b) After discussion, citizens of France become more critical of the United States and its intentions with respect to economic aid.[44] (c) After discussion, whites predisposed to show racial prejudice offer more negative responses to the question whether white racism is responsible for conditions faced by African-Americans in American cities.[45] (d) After discussion, whites predisposed not to show racial prejudice offer more positive responses to the same question.[46] As statistical regularities, it should follow, for example, that that those moderately critical of an ongoing war effort will, after discussion, sharply oppose the war; that a group moderately predisposed to hire a certain job candidate will, after discussion, support the application with considerable enthusiasm; that people tending to believe in the inferiority of a certain racial group will be entrenched in this belief as a result of discussion.

The phenomenon of group polarization has conspicuous importance to the operation of deliberating bodies of relevance to law and politics, including legislatures, commissions, multimember courts, and juries. I will return to this point shortly; for now notice a few obvious possibilities. Members of a political party, or of the principal political parties, may polarize as a result of internal discussions; party-line voting is sometimes explicable partly on this ground. A set of judges with similar predilections on a three-judge panel may well produce a more extreme ruling than any individual member would write if he were judging on his own. Extremist groups will often become more extreme; as we will soon see, the largest group polarization typically occurs with individuals already inclined toward extremes. With respect to deliberating juries, a recent study[47] found significant group polarization with respect to "numerical punishment ratings" on a bounded numerical scale. For high punishment ratings, groups tended to generate numbers <u>higher</u> than the median of individual predeliberation judgments; for low punishment ratings, groups tended to generate numbers <u>lower</u> than the median of individual predeliberation judgments. This is precisely the pattern that group polarization would predict.

B. Risky Shifts and Cautious Shifts

Group polarization was first found in a series of experiments involving risk-taking decisions.[48] Before 1961, conventional wisdom had been that as compared with the individuals who compose it, a group of decision-makers – for example a committee or board – would be

---

[41] Of course, when different deliberating groups polarize in different directions, the consequence can be great among-group variance.

[42] These include the United States, Canada, Germany, and France. Of course it is possible that some cultures would show a greater or lesser tendency toward polarization; this would be an extremely interesting area for empirical study.

[43] See D.G. Myers, Discussion-Induced Attitude Polarization, 28 Human Relations 699 (1975).

[44] Brown, Social Psychology 224 (2d ed. 1983).

[45] D.G. Myers and G.D. Bishop, The Enhancement of Dominant Attitudes in Group Discuission, 20 J Personality and Soc. Psych. 286 (1976),

[46] See id.

[47] See David Schakde et al., Are Juries More Erratic Than Individuals?, Colum L Rev (2000).

[48] I draw in this and the following paragraph on Brown, supra note, at 200-206.

likely to favor a compromise and thus to avoid risks. But the relevant experiments, originally conducted by Stoner, found otherwise; they identified what has become known as the "risky shift."[49] Deliberation tended to shift group members in the direction of greater risk-taking; and deliberating groups, asked to reach a unanimous decision, were generally more risk-inclined – sometimes far more risk-inclined – than the mean individual member, predeliberation.

In the original experiments, male graduate students of industrial management were asked a range of questions involving risk: whether someone should choose a safe or risky play in the last seconds of a football game; whether someone should invest money in a low-return, high-security stock or instead a high-return, lower security stock; whether someone should choose a high prestige graduate program in which a number of people fail to graduate or a lower prestige school where everyone graduates. In one problem, for example, people were asked to say whether a person now having a secure, lifetime job should take a new job, with a new company with an uncertain future. People were asked about the lowest probability of "financial soundness" that would justify the person with the secure job from taking the new position. In Stoner's studies, people first studied the problems – twelve total – and recorded an initial judgment; they were then asked to reach a unanimous decision as a group. People were finally asked to state their private judgments after the group judgment had been made; they were informed that it was acceptable for the private judgment to differ from the group judgment.

For twelve of the thirteen groups, the group decisions showed a repeated pattern toward greater risk-taking -- that is, after discussion, the unanimous outcome tended to assess the necessary likelihood of financial soundness as consistently lower than the median judgment of the group predeliberation. In addition, there was a clear shift toward greater risk-taking in private opinions as well. Only 16% were moved toward greater caution; 45% did not change at all; and a full 39% moved in the direction of greater risk-taking. This shift – the "risky" shift – was promptly duplicated in a number of diverse studies, some involving all men and some involving all women.

We should distinguish at this point between two aspects of these findings, not always separated in the psychological literature and both of relevance to law and policy. The first involves the movement of deliberating groups, for whom a group decision is necessary, toward the group's extreme end; call this (inelegantly) group polarization toward within-group extremes. This means that if a group decision is required, the group will tend toward an extreme point, given the original distribution of individual views. Undoubtedly the group's decision rule will matter here; a requirement of unanimity may well, for example, produce a shift toward the most extreme points, at least if those with the most extreme views are least tractable and most confident. The second involves the movement of (even private) individual judgments as a result of group influence; call this individual polarization toward within-group extremes. This means that to the extent that private judgments are moved by discussion, it will be toward a more extreme point in the direction set by the original distribution of views.

---

[49] See J.A.F. Stoner, A Comparison of Individual and Group Decisions Including Risk, unpublished master's thesis, School of Management, Massachusetts Institute of Technology; J.A.F. Stoner, Risky and Cautious Shifts in Group Decisions, 4 J Experimental Social Psych. 442 (1968).

A possible (and contemporaneous) reading of Stoner's early studies would be that group dynamics are such as to move people – both groups and individuals within them -- in the direction of greater risk-taking. But this conclusion would be much too simple. Later studies showed that under certain conditions, it was possible, even easy to induce a "cautious shift" as well. Indeed, certain problems reliably produced cautious shifts. The principal examples involved the decision whether to marry and the decision whether to board a plane despite severe abdominal pain possibly requiring medical attention. In these cases, deliberating groups moved toward caution, as did the members who composed them.

As yet there is no simple account of what kinds of problems will produce what kinds of shifts; but the identification of risky and cautious shifts has helped produce a general account of how much, and in what direction, people will tend to move. In Stoner's original data, subsequent researchers noticed, the largest risky shifts could be found when group members "had a quite extreme risky initial position," in the sense that the predeliberation votes were weighted toward the risky end, whereas the items "that shifted a little or not at all started out near the middle of the scale."[50] Thus the direction of the shift seemed to turn on the location of the original disposition, and the size of the shift depended on the extremeness of that original disposition. A group of very cautious individuals would produce a significant shift toward greater caution; a group of individuals inclined toward risk-taking would produce a significant shift toward greater risk-taking; and groups of individuals in the middle would produce smaller shifts in the direction indicated by their original disposition. In short, "group discussion moves decisions to more extreme points in the direction of the original inclination . . . , which means shift to either risk or caution in the direction of the original disposition, and the size of the shift increases with the degree of the initial polarization."[51] Similar results have been found in many contexts, involving, for example, questions about economic aid, architecture, political leaders, race, feminism, and judgments of guilt or innocence.[52] Polarization has been found for questions of obscure fact (eg, how far Sodom on the Dead Sea is below sea level) as well as for evaluative questions, including political and legal issues[53] and even the attractiveness of people in slides.[54]

IV. Mechanisms

A. Two Mechanisms

What explains group polarization? It is tempting to think that conformity plays a large role, and as the Asch experiments suggest, individual judgments have been found to be greatly influenced by the desire to conform. Perhaps conformity is sometimes at work, but group polarization is not a matter of conformity; people do not shift to the mean of initial positions. The relevant movement goes to one or another side. Indeed, this is what defines, and what is most interesting about, group polarization.

---

[50] Brown at 211.

[51] Brown, supra, at 211.

[52] See id.

[53] A relatively recent treatment is Russell Spears, Martin Lee, and Stephen Lee, De-Individuation and Group Polarization in Computer-Mediated Communication, 29 British J Soc Psych 121 (1990

[54] Turner et al., supra, at 153.

There have been two main explanations for group polarization, both of which have been extensively investigated.[55] Massive support has been found on behalf of both explanations.[56]

1. <u>Social comparison</u>. The first, involving social comparison, begins with the claim that people want to be perceived favorably by other group members, and also to perceive themselves favorably. Once they hear what others believe, they adjust their positions in the direction of the dominant position. They may want to signal, for example, that they are not cowardly or cautious, and hence they will frame their position so that they do not appear such by comparison to other group members.[57] With respect to risk-taking activity, people want to occupy a certain position in comparison to others, and before they hear what other people think, they assume that they do in fact occupy that position. But when they hear what other people think, they find, often, that they occupy a somewhat different position, and they shift accordingly. The result is to press the group's position toward one or another extreme, and also to induce shifts in individual members. The same appears to happen in other contexts. People may wish, for example, not to seem too enthusiastic, or too restrained in their enthusiasm for, affirmative action, feminism, or an increase in national defense; hence their views may shift when they see what other group members think. The result will be both group and individual polarization toward within-group extremes.

The dynamic behind the social comparison explanation is that most people may want to take a position of a certain socially preferred sort – in the case of risk-taking, for example, they may want to be perceived (and to perceive themselves) as moderate risk-takers, and their choice of position is partly a product of this desire.[58] No one can know what such a position would be until the positions of others are revealed.[59] Thus individuals move their judgments in order to preserve their image to others and their image to themselves. A key claim here is that information alone about the actual positions of others – without discussion -- will produce a shift. Evidence has confirmed this fact; mere exposure induces a substantial risky shift (though it is less substantial than what is produced by discussion – about half as large).[60] This effect helps explain a shift toward caution (the "cautious shift") as well.[61] While highly suggestive, the "mere

---

[55] Isenberg, supra, and Brown, supra, review this literature; see also Turner et al., supra, at 142-70, for an overview and an attempt to generate a new synthesis.

[56] Note that conformity does not explain group polarization. People are not attempting to conform, even under the social comparison theory; they are attempting to maintain their relative position, and the revelation of the views of others shifts people's conception of what judgment is necessary to maintain that position. See Myers, supra note, at 562, indicating that people "want to perceive themselves as somewhat different from others" and that "people want to differentiate themselves from others, to a small extent and in the right direction."

[57] On signalling generally, see Eric Posner, Symbols, Signals, and the Law (forthcoming 2000).

[58] For a quite vivid demonstration of such a process in the enactment of the Clean Air Act, one that does not, however, identify the mechanisms, discussed here, see Bruce Ackerman, John Millian, and Donald Elliott, Toward a Theory of Statutory Evolution: The Federalization of Environmental Law, 1 J. L. Econ. & Organization 313 (1985).

[59] "Once the real locations of the mean was known, should it not be the case, granting that everyone wanted to see himself as reasonably audacious, that those who were really below the mean would be motivated to adopt riskier positions and so change the mean and produce the risky shift?" Brown, supra, at 214.

[60] Teger and Pruitt (1967).

[61] Investigations of social influence have emphasized both one-upmanship and the removal of pluralistic ignorance, that is, ignorance of what other people think (or are willing to say they think). Note that it is implicit in these findings that people seem to want not to conform, but to be different from others in a desirable way. "To be virtuous . . . is to be different from the mean – in the right direction and to the right degree." Brown, supra note, at 469.

exposure" finding does not confirm the social influence account; it is possible that the views of others simply provide an informational signal, quite apart from arguments, and hence that people move not in order to maintain reputation, but to do what is right. (Recall the discussion of informational cascades.)

The social influence explanation invokes factors similar to those that underlie the reputational cascade. A major difference is that the social influence explanation concerns presentation to self as well as presentation to others. Note also that group polarization may or may not be a result of any cascade effect; the question is whether the accumulation of views from others operates in the form of a cascade. Existing work on group polarization does not answer this question.

2. <u>Persuasive arguments</u>. The second explanation, emphasizing the role of persuasive arguments, is based on a common sense intuition: that any individual's position on an issue is partly a function of which arguments presented within the group seem convincing. The choice therefore moves in the direction of the most persuasive position defended by the group, taken as a collectivity. Because a group whose members are already inclined in a certain direction will have a disproportionate number of arguments supporting that same direction, the result of discussion will be to move individuals further in the direction of their initial inclinations. The key is the existence of a limited argument pool, one that is skewed (speaking purely descriptively) in a particular direction.

The persuasive arguments theory begins with the suggestion that if a group is deliberating about some difficult question with a factual answer (how many countries are there in Africa, for example, or how many people were on the planet in 1900), discussion will typically produce some movement, not toward the mean, but toward the minority view on which one or a few members have accurate information. There is, moreover, empirical evidence that with respect to facts, deliberation produces movements toward accuracy.[62] Of course many of the questions involving group polarization do not have purely factual answers. But a key aspect of those discussions is that the person with the correct answer is likely to state his view with a high degree of confidence, and also be able to make some argument in favor of that view. Novel arguments, bringing up fresh points, are especially likely to be persuasive. In any case members of a group will have thought of some, but not all, of the arguments that justify their initial inclination; consider the question whether to take risks or to be cautious. In discussion, arguments of a large number of individuals are stated and heard, but the total argument pool will be tilted in one or another direction, depending on the predispositions of the people who compose the group; hence there will be a shift in the direction of the original tilt.[63]

When people hear arguments that they perceive as valid, or find to be memorable, vivid, new, or weighty simply by virtue of emphasis and repetition, they will shift in the direction suggested by those arguments. If a group of moderately feminist women becomes more feminist, a group moderately opposed to gun control more extremely so, and so forth, one reason is that the argument pool of any such group will contain a preponderance of arguments in the direction

---

[62] See James Fishkin and Robert Luskin, Bringing Deliberation to the Democratic Dialogue, in The Poll With A Human Face 3, 29-31 (Maxwell McCombs and Amy Reynolds eds. 1999).
[63] Brown, supra, at 219.

suggested. The suggestion is that group polarization will occur when convincing arguments produce a shift in the direction of prediscussion inclinations, revealed in the means of the initial decisions

There is an obvious analogy here to the informational cascade. In fact we can safely assume that group polarization sometimes occurs via a kind of informational cascade, as the statements of particular people begin a cascade process that culminates in extremism. The difference is that for cascade effects, what is crucial is the very fact of the belief, not its grounds, whereas for persuasive arguments to work, what is crucial is that arguments be offered and be found persuasive. It is also unclear whether any particular group polarization involves cascade effects at all; undoubtedly what sometimes happens is not a cascade effect, in which a large number of people successively "fall," but a simple accumulation of arguments, eventually imposing weight on people whose views are subject to change.

B. Refinements -- and Depolarization

These are statistical regularities, no more. Of course not all groups polarize; some groups end up in the middle, not toward either extreme. Note that in Stoner's original experiments, one of the twelve deliberating groups showed no polarization at all. Nor is it hard to understand why this might be so. If the people defending the original tendency are particularly unpersuasive, group polarization is unlikely to occur. If the outliers are especially convincing, groups may even shift away from their original tendency and in the direction held by few or even one.[64] In addition, affective factors appear to be quite important and complementary to persuasive arguments. People are less likely to shift if the direction advocated is being pushed by unfriendly group members; the chance of shift is increased when people perceive fellow members as friendly, likeable, and similar to them.[65] Physical spacing tends to reduce polarization; a sense of common fate and intragroup similarity tend to increase it, as does the introduction of a rival "outgroup."[66] Part of the reason for group polarization appears to be that as a class, extreme positions tend to be less tractable and more confidently held. This point is an important complement to the persuasive arguments theory[67]: The persuasiveness of arguments depends, not surprisingly, not simply on the grounds given, but also on the confidence with which they are

---

[64] This is of course the theme of the movie Twelve Angry Men, where the single hold-out, played by Henry Fonda, shifts the judgment of the jury.

[65] See Hans Brandstatter,, Social Emotions in Discussion Groups, in Dynamics of Group Decisions (Hans Brandstatter et al. ed. 1978). Turner et al., supra note, at 154-59, attempt to use this evidence as a basis for a new synthesis, one that they call "a self-categorization theory of group polarization," id. at 154. In this account, "persuasion is dependent upon self-categorizations which create a common identity within a group," and polarization occurs "because group members adjust their opinion in line with their image of the group position (conform) and more extreme, already polarized, prototypical responses determine this image." Id. at 156. The key point here is that when a group is tending in a certain direction, the perceived "prototype" is determined by where the group is leaning, and this is where individuals will shift. Id. at 156. As Turner et al. suggest, their account shows "overlap with many aspects of social comparison and persuasive arguments models," and because of the overlap, I do not discuss it as a separate account here. For possible differences in predictions, and supporting evidence, see id. at 158-70. An especially interesting implication, perhaps in some tension with the persuasive arguments theory, is that a group of comparative extremists will show a comparatively greater shift toward extremism. See id. at 158.

[66] See Turner et al., supra note, at 153.

[67] See Maryla Zaleska, The Stability of Extreme and Moderate Responses in Different Situations, in Group Decision Making, supra, at 163, 164.

articulated. (Consider here both juries and multimember courts.) Group polarization can also be fortified through "exit," as members leave the group because they reject the direction in which things are heading. If exit is pervasive, the tendency to extremism can be greatly aggravated.

Notably, the persuasive arguments theory implies that there will be "depolarization," or convergence toward the middle, if and when new persuasive arguments are offered that are opposite to the direction initially favored by group members. There is evidence for this phenomenon as well. [68] Depolarization, rather than polarization, will also be found when the relevant group consists of individuals drawn equally from two extremes (a point to which I will return).[69] Thus if people who initially favor caution are put together with people who initially favor risk-taking, the group judgment will move toward the middle.

Group members with extreme positions generally change little as a result of discussion or shift to a more moderate position.[70] Consider a study[71] consisting of six-member groups specifically designed to contain two subgroups (of three persons each) initially committed to opposed extremes; the effect of discussion was to produce movement toward the center. One reason may be the existence of partially shared persuasive arguments in both directions.[72] Interestingly, this study of opposed subgroups found the greatest depolarization with obscure matters of fact (e.g., the population of the United States in 1900) -- and the least depolarization with highly visible public questions (e.g., whether capital punishment is justified). Matters of personal taste depolarized a moderate amount (e.g., preference for basketball or football, or for colors for painting a room).[73]

These findings fit well with the persuasive arguments account of polarization. When people have a fixed view of some highly salient public issue, they are likely to have heard a wide range of arguments in various directions, producing a full argument pool, and an additional discussion is not likely to produce movement. Hence "familiar and long-debated issues do not depolarize easily."[74] With respect to such issues, people are simply less likely to shift at all.

It also matters whether people think of themselves, antecedently or otherwise, as part of a group, with a degree of solidarity. If they think of themselves in this way, group polarization is all the more likely, and it is likely too to be more extreme.[75] Thus when people are "de-

---

[68] A third possibility is that hearing other similar opinions produces greater confidence in individual positions, opening members to a more extreme judgment in the same direction. raised recently by Heath and Gonzales . See Chip Heath and Richard Gonzales, Interaction With Others Increases Decision Confidence But Not Decision Quality: Evidence Against Information Collection Views Of Interactive Decision Making, 61 Organizational Behavior and Human Decision Processes 305-326 (1997).

[69] See H. Burnstein, Persuasion As Argument Processing, in Group Decision Making (H. Brandstetter, J.H. Davis, and G. Stocker-Kreichgauer eds., 1982).

[70] Ferguson and Vidmar, Effects of Group Discussion on Estimates of Risk Levels, 20 J Pers and Social Psych 436 (1971).

[71] Brown, supra, at 225.

[72] Amiram Vinokur and Eugene Burnstein, The Effects of Partially Shared Persuasive Arguments on Group-Induced Shifts, 29 J Personality & Social Psych. 305 (1974).

[73] Amiram Vinokur and Eugene Burnstein, supra, at 884.

[74] Brown, supra, at 226.

[75] See Russell Spears, Martin Lee, and Stephen Lee, De-Individuation and Group Polarization in Computer-Mediated Communication, 29 British J Soc Psych 121 (1990); Patricia Wallace, The Psychology of the Net (1999).

individualized," in the sense that the context emphasizes each person's membership in the social group engaging in deliberation, polarization increases.[76]

An especially interesting experiment attempted to investigate this point by manipulating two variables.[77] First, some subjects were "de-individualized" by having to work on computers in separate rooms, whereas others were asked to work in a single office with desks facing each others (the "individualized" condition), In the de-individualized condition, visual anonymity was increased. Second, some subjects were given instructions in which group membership was made salient (the "group immersion" condition), whereas others were not (the "individual" condition). For example, subjects in the group immersion conditions were told that their group consisted solely of first-year psychology students, and that they were being tested as group members rather than as individuals. All conditions were held constant in one respect: Every subject was told that people like them tended to support one or another view. The relevant issues involved affirmative action, government subsidies for the theatre, privatization of nationalized industries, and phasing out nuclear power plans.

The results were quite striking. There was the least group polarization in the de-individuated-individual condition; group polarization was greatest in the de-individuated/group immersion condition, when group members met relatively anonymously and when group identity was emphasized. Interestingly, there was no significant difference in polarization between the two individuated conditions (with and without emphasis on group immersion). From this experiment, it is reasonable to speculate that polarization is most likely to occur, and to be most extreme, under circumstances in which group membership is made salient and people have a high degree of anonymity. There is obviously a potential lesson here about the effects of group deliberation on the Internet,[78] a point to which I will return.

These remarks suggest some general, common-sensical conclusions about how and when group discussion will move predeliberation opinions. Views based on a great deal of thought are least likely to shift; depolarization can occur with equal subgroups tending in opposite directions; groups will usually shift in the direction of an accurate factual judgment where one or more members knows the truth; where views are not firmly held, but where there is an initial predisposition, group polarization is the general rule. Undoubtedly generalizations of this sort bear on shifts in individual views among many deliberating bodies.

B. Iterated "Polarization Games"?

The logic of group polarization suggests that if participants engage in repeated discussions – if, for example, they meet each month, express views, and take votes – there should be repeated shifts toward, and past, the defined pole. Thus, for example, if a group of people is thinking about genetic engineering of food, or the minimum wage, or the World Trade Organization the consequence of their discussions, over time, should be to lead in quite extreme directions. In these iterated "polarization games," deliberation over time should produce a

---

[76] Spears et al., supra, at 122-124.
[77] Id.
[78] See Patricia Wallace, The Psychology of the Internet 73-79 (1999).

situation in which individuals hold positions more extreme than those of any individual member before the series of deliberations began.

This is only a thought experiment; there appears to be no study of such iterated polarization games. But the hypothesized outcome is less fanciful than it might seem. In the jury study referred to above, deliberating groups frequently came up with punishment ratings, and with dollar awards, as high as or even higher than that of any individual, pre-deliberation.[79] And it is not difficult to think of real-world groups in which the consequence of deliberation, over time, appears to be to shift both groups and individuals to positions that early on, they could not possibly have accepted.[80] Iterated polarization games seem to be an important real-world phenomenon. But this raises two questions: Why and when do groups stop polarizing? Why and when do they end up at a certain point, or even shift in the opposite direction? Nothing in the literature on group polarization adequately answers these questions.[81] But it is possible to speculate that polarization often ends or reverses as a result of some external shock – as, for example, when new members add new arguments, or when the simple self-interest of political leaders produces a shift in direction,[82] or when new circumstances, of fact or value, alter the perspectives and incentives of group members. Social cascades often change direction as a result of such external shocks, as through the release of new information[83]; the same processes seem to terminate or to reverse group polarization.

C. A Wrinkle: "Rhetorical Asymmetry"

Interestingly – and in a noteworthy qualification of the general literature on group polarization – the previously discussed study of punitive damage awards by juries found a striking pattern for dollar awards.[84] For any dollar award above zero, the general effect of deliberation was to increase awards above those of the median voter. Dollar awards did not simply polarize; while higher awards increased dramatically, as compared to the median of predeliberation votes, low awards increased as well. Why is this?

Both the original experiment and a follow-up experiment suggest that this result is a product of a "rhetorical asymmetry" that favors, other things being equal and in any contest, the person or persons urging higher awards. Thus the fact of systematic increases in dollar awards in strongly suggested of a general tendency toward upward movement; a subsequent experiment, limited to University of Chicago law students, confirmed this effect, with a substantial majority of subjects agreeing that it was easier, other things being equal, to argue for higher awards than

---

[79] David Schkade, Cass R. Sunstein, and Daniel Kahneman, Are Juries More Erratic Than Individuals?, 100 Colum L Rev (forthcoming 2000).

[80] For a relevant discussion of deliberating groups in the 1960s, see James Miller, Democracy Is In The Streets (1993); see also Timur Kuran, Ethnic Norms and Their Transformation Through Reputational Cascades, 27 J Legal Stud 623, 648 (1998).

[81] See Turner et al., supra, at 152, suggesting that there is no clear answer to the question for "what range of situations" polarization is predicted.

[82] See the discussion of legislators below; and note, as inevitably noisy real-world examples, the centrist shift of the Democratic Party in 199101992, and the apparent centrist shift of the Republican Party in 1999-2000.

[83] See David Hirschleifer, The Blind Leading The Blind, in The New Economics Of Human Behavior (1997).

[84] See David Schkade, Cass R. Sunstein, and Daniel Kahneman, Are Juries More Erratic Than Individuals?, 100 Colum L Rev (forthcoming 2000).

for lower ones.[85] In our culture, and in light of existing norms, the person favoring the higher amount for punitive damages appears likely to be more convincing than the person favoring the lower amount. It is important to emphasize that this asymmetry operates independently of any facts about the individual case. The reason appears to be that with respect to dollar awards involving a corporate defendant, stronger arguments – "we need to deter this kind of conduct," "we need to send a powerful signal," "we need to attract their attention" – tend to have comparatively greater weight.

Undoubtedly there are many other contexts containing rhetorical asymmetry, and undoubtedly the asymmetry can affect outcomes, as it did in the jury study. A great deal of empirical work remains to be done on this question; too little is known to say why and when such an asymmetry is at work. Existing cultural norms are the underlying source of any rhetorical asymmetry, and such norms vary over space and time, producing shifts in rhetorical asymmetry. In any case it is not difficult to generate possible examples. Legislative judgments about criminal punishment may, for example, involve an asymmetry of exactly this kind. In certain settings, those favoring lower taxes, or more aid for scholarship students, or greater funding for environmental protection may have a similar rhetorical advantage. Much remains to be explored. For present purposes the point is that group polarization may be aggravated or attenuated if one or another side has a systematic advantage in rhetoric. Perhaps the most striking implication is that when there is an initial distribution of views in a certain direction, and when a more extreme movement in that direction has a rhetorical advantage, quite extreme shifts can be expected.[86]

D. Is Group Polarization Rational?

In both economics and law, a great deal of attention has recently been paid to the question whether human beings are "rational," or "quasi-rational," or subject to irrationality.[87] There is an obvious question whether the phenomenon of group polarization raises doubts about rational actor models in economics or law. The answer is that for the most part, individual behavior within groups, as described thus far, creates no such doubts. It is certainly rational to make assessments on the basis of arguments offered; if the most numerous and convincing arguments seem to justify a shift, individual shifts are entirely rational.[88] More difficult questions might seem to be raised by "social influence" accounts of group polarization. But it is certainly rational for people to care about their reputations. If they are changing their assessment because of reputational considerations, what must be said is that maintaining a certain reputation is part of what people care about (and there is nothing irrational about that). If people shift not for reputational reasons but because of a certain self-conception – if, for example, they think of themselves as people who are bold, or committed to a strong national defense, or left of center on

---

[85] See id at XX.

[86] Data from id. strongly support this speculation, with many deliberating juries producing dollar awards higher, and sometimes significantly higher, than that of the higher individual pre-deliberation. Data to be added.

[87] Within economics, see, e.g., Gary Becker, Accounting For Tastes (1998); Richard Thaler, Quasi-Rational Economics (1993). Within law, see Symposium, The Legal Implications of Psychology, 51 Vand. L. Rev. 1499 (1998); Behavioral Law and Economics (Cass R. Sunstein ed. 2000).

[88] Compare Hirshleifer, supra note (emphasizing the rationality of participation in informational cascades).

issues of race – a change in position, after exposure to the views of others, also seems entirely rational.

The problem with group polarization is not that people subject to it suffer from some cognitive or motivational defect. The problem is instead that people may be shifted, as a result of entirely rational processes, in the direction of factual, legal, or moral mistakes.

E. The Relativity of Polarization and Polarization Framing

Notwithstanding the rationality of those subject to polarization, it should be emphasized that in laboratory studies, polarization occurs in terms of a specified issue and a specified scale. The issue for exploration is therefore framed in a certain way, and here there is a potential for manipulation. The same group of individuals, for example, might be inclined to be supportive of greater employment opportunities for women and also inclined to be skeptical about "feminism"; and polarization could drive otherwise identical groups toward more extreme positions on both questions, so much so that, in theory, one group could become very strongly committed to women's employment opportunities that it embraced feminism, whereas another group could become so skeptical of feminism that it raised questions about greater employment opportunities for women. Here there is a lesson about the pervasive importance of "framing" in generating positions about disputed questions.[89] But there is a twist on the conventional view: In the presence of polarization, questions can be framed in such a way as to shift groups, and individuals who constitute them, in distinctive and even inconsistent directions.

Now if people attempt to square their various judgments with one another, in an attempt to reach reflective equilibrium,[90] inconsistent shifts are less likely, and people ought to be less vulnerable to framing effects. In the real world, however, it is likely that polarization occurs around issues as socially framed; cultural movements of various sorts – toward greater ethnic identification, in favor of stronger national defense, on behalf of taxpayer support for the arts – are a likely consequence. Undoubtedly political entrepreneurs, with self-interested or altruistic agendas, are in some sense aware of this fact, and attempt to produce shifts along the scale that has been made salient.

V. Implications and Illustrations

A. Outside the Laboratory

Group polarization should have a large effect on any deliberating group or institution; its effects are hardly limited to the laboratory. Religious organizations tend, for example, to strengthen group members' religious convictions, simply by virtue of the fact that like-minded people are talking to one another.[91] Indeed religious groups amplify the religious impulse, especially if group members are insulated from other groups; the result can be to lead people in

---

[89] See Donald A. Redermeier et al., Understanding Patients' Decisions, 270 JAMA 72, 73 (1993).
[90] See John Rawls. A Theory of Justice (1971).
[91] See David G. Myers, Polarizing Effects of Social Interaction, in Group Decision Making 125, 137-38 (Hermann Barndstatter et al. eds 1982).

quite bizarre directions.[92] Survey evidence shows that dramatic social events, like the assassination of Martin Luther King and civil rights disturbances, tend to polarize attitudes, with both positive and negative attitudes increasing within demographic groups.[93] A similar process can harden attitudes toward outsiders and social change; thus proposals "for establishment of a halfway house or a correctional facility have typically elicited private apprehensions which, after discussion, become polarized into overt paranoia and hostility."[94]

B. Outgroups

Group polarization has particular implications for insulated "outgroups." Recall that polarization increases when the deliberating group is able to define itself by contrast to some other contrasting group[95]; outgroups are in this position – of self-contrast to others – by definition. Excluded by choice or coercion from discussion with others, such groups may become polarized in quite extreme directions, often for no better reason than group polarization. Extremism on the part of such groups (not excluding murders and suicides) is a possible result,[96] especially if we consider the fact that extreme groups show comparatively greater polarization.[97] There is also likely to be some rhetorical asymmetry within such groups, so that arguments in a certain directions have the automatic upper hand. Consider, for example, a group of people who tend to believe that academic freedom is threatened by the tendency to "political correctness" in university life; in a debate about how much attention should be paid to (say) gender studies in the curriculum, skeptics are likely to have the upper hand.

The tendency toward polarization among outgroups raises some doubts about the idea that certain group discussion produce "consciousness raising." It is possible, at least, that the consequence of discussion is not to raise consciousness (an ambiguous term to be sure), but to produce group polarization in one direction or another -- and at the same time to increase confidence in the position that has newly emerged.[98] This does not mean that consciousness is never raised; undoubtedly group discussion can identify and clarify problems that were previously repressed, or understood as an individual rather than social product. But nothing of this sort is established by the mere fact that views have changed and coalesced, and are held, post-discussion, with a high degree of confidence.[99]

C. Feuds, Ethnic and International Strife, and War

Some of the relevant processes are at work in feuds of all kinds; one of the characteristic features of feuds is that the feuding groups tend to talk only to one another, fueling and

---

[92] See Leon Festinger et al., When Prophecy Fails (1956).

[93] See R.T Riley and T.F. Pettigrew, Dramatic Events and Attitude Change, 34 J Personality and Social Psych 1004 (1976).

[94] Myers, supra, at 135.

[95] See Turner et al., supra note, at 151.

[96] Cf. Aronson, supra note (discussing mass suicide at Jones Beach).

[97] See supra; Turner et al., supra, at 158, 167-70.

[98] See Chip Heath and Rich Gonzalez, Interaction With Others Increases Decision Confidence But Not Decision Quality: Evidence Against Information Collection Views of Interactive Decision Making, 61 Organizational Behavior and Human Decision Processes 305 (1995).

[99] See id.

amplifying their outrage, and solidifying their impression of the relevant events.[100] Informational and reputational forces are very much at work here, producing cascade effects, and group polarization sometimes leads members to increasingly extreme positions. It is not too much of a leap to suggest that these effects are also present within ethnic groups and even nations, notwithstanding the obvious fact that here there is a high degree of heterogeneity, and deliberation cannot occur among all members at the same time.

Timur Kuran, for example, has explored the phenomenon of "ethnification." Kuran's basic claim is that in many nations, including Turkey and the former Yugoslavia, ethnic strife is not a reawakening of long-suppressed resentments, but instead a product of reputational cascades. In this process, a failure to engage in ethnically identified activity produces reputational sanctions, which grow in intensity over time, as increasing numbers of people join the cascade. Hence "the fears and antagonisms that accompany high levels of ethnic activity may be a result of ethnification rather than its root cause."[101] Kuran does not refer to group polarization. But an understanding of this phenomenon would much fortify his analysis, by showing how within-group discussion (which is, under conditions of ethnification, an increasingly large percentage of total discussion) can ensure that ethnic groups, and individual members of ethnic groups, end up with a far stronger ethnic identification than the median member, before discussions began. In the extreme case, the result might be war.[102] And when a war begins, group polarization, if it operates at the national level, can help ensure continued hostility and antagonism.

D. The Internet and Mass Deliberation

Many people have expressed concern about processes of social influence on the Internet.[103] The general problem is said to be one of fragmentation, with certain people hearing more and louder versions of their own preexisting commitments, thus reducing the benefits that come from exposure to competing views and unnoticed problems.[104] But an understanding of group polarization heightens these concerns and raises new ones. A "plausible hypothesis is that the Internet-like setting is most likely to create a strong tendency toward group polarization when the members of the group feel some sense of group identity."[105] If certain people are deliberating with many like-minded others, views will not be reinforced but instead shifted to more extreme points. This cannot be said to be bad by itself – perhaps the increased extremism is good – but it is certainly troublesome if diverse social groups are led, through predictable mechanisms, toward increasingly opposing and ever more extreme views. It is likely that processes of this general sort have threatened both peace and stability in some nations[106]; while dire consequences are unlikely in the United States, both fragmentation and violence are predictable results. As we have seen, group polarization is intensified if people are speaking anonymously and if attention is drawn,

---

[100] See Jon Elster, The Cement of Society (1994).
[101] See Timur Kuran, Ethnic Norms and Their Transformation Through Reputational Cascades, 27 J Legal Stud 623, 648 (1998).
[102] See id. at 650-51.
[103] See Lawrence Lessig, Code and Other Laws of Cyberspace 186 (1999); Andrew Shapiro, The Control Revolution (1999).
[104] See id.
[105] See Patricia Wallace,, supra note, at 73-84.
[106] See Timur Kuran, supra note, at 635-650.

through one or another means, to group membership. Many Internet discussion groups have precisely this feature. It is therefore plausible to speculate that the Internet may be serving, for many, as a breeding group for extremism.

Consider in this regard a study not of extremism, but of serious errors within working groups, both face-to-face and online.[107] The purpose of the study was to see how groups might collaborate to make personnel decisions. Resumes for three candidates, applying for a marketing manager position, were placed before the groups; the attributes of the candidates were rigged by the experimenters so that one applicant was best matched for the job described. Packets of information were given to subjects, each containing only a subset of information from the resumes, so that each group member had only part of the relevant information. The groups consisted of three people, some operating face-to-face, some operating on-line. Two results were especially striking: Group polarization was common; and almost none (!) of the deliberating groups made what was conspicuously the right choice, because they failed to share information in a way that would permit the group to make an objective decision. In on-line groups, the level of bias was especially high, in the sense that members tended to share positive information about the winning candidate and negative information about the losers, while also suppressing negative information about the winner and positive information about the losers. These contributions served to "reinforce the march toward group consensus rather than add complications and fuel debate,"[108] This tendency was twice at large within the online groups.

It is much too early to offer a confident account of the consequences of group deliberation via computer and on the Internet. But what has been said thus far should be sufficient to show that group polarization may be especially pronounced under conditions of anonymity, in a way that magnifies mistakes and biases. Though the study just described did not involve political or moral issues, the results are plausibly taken to suggest that one-sidedness, and consequently extremeness, can be heightened when communication occurs via computer.

E. Legal and Political Institutions

With respect to legal and political institutions, there is generally little direct evidence; but it is possible to venture several points.

1. Juries. Group polarization is well-documented on juries; this is the only legal institution for which direct evidence exists. In experimental settings, polarization has been found in numerous settings with respect to guilt and innocence, and indeed this appears to be an uncontradicted finding.[109] Outside of the experimental setting, we know that the predeliberation verdict predicts the final outcome 90% of the time, in cases where juries do not hang; this provides "powerful presumptive evidence that group polarization occurs in real juries."[110]

As noted, a more recent study of 300 deliberating juries found massive group polarization with respect to bounded punishment scales; groups whose median pre-deliberation vote was 3 or

---

[107] See Hightower and Sayeed, supra note.
[108] Wallace, supra note, at 82.
[109] See Brown, supra note, at 227-29 (collecting studies).
[110] Id. at 239.

less tended to generate verdicts below that of the median voter, whereas groups whose median pre-deliberation vote was above 3 tended to generate verdicts above that of the median voter. Indeed, many such juries ended up with verdicts as low or lower (for the low verdicts) as that of the lowest predeliberation voter, and as high or higher (for the high verdicts) as that of the highest predeliberation voter. I have also noted that with respect to dollars, this study did not find group polarization, at least in any simple form; positive dollar amounts generally increased, because of the rhetorical asymmetry referred to above. On the other hand: As compared to the median of predeliberation judgments, dollar amounts increased far more at the high end, and this effect is broadly consistent with the idea of group polarization.

2. Independent regulatory commissions. The twentieth century has seen the rise of a number of "independent" regulatory commissions, including the Federal Communications Commission, the Federal Trade Commission, and the National Labor Relations Board. These commissions have attracted attention mostly because their members are immune from plenary removal power by the President.[111] An equally striking but generally overlooked provision of the relevant statutes requires bipartisan membership: The independent commissions must be divided between Republicans and Democrats. A simple and undoubtedly correct explanation of this unusual requirement is that Congress wanted to ensure that no commission would be dominated by any single party. But an understanding of group polarization would strengthen any such concern on Congress' part. An independent agency – the FCC, the NLRB, the CPSC -- that is all-Democratic, or all-Republican, might polarize toward an extreme position, likely more extreme than that of the median Democrat or Republican, and possibly more extreme than that of any member standing alone. A requirement of bipartisan membership can operate as a check against movements of this kind.

3. Multimember courts. Group polarization should also occur on multimember courts. Notwithstanding platitudes about judicial neutrality, judges often have a great deal of latitude, sometimes in the ultimate outcome, more often in determining the reach of their decision. If a court consists of three or more like-minded judges, it may well end up with a relatively extreme position, more extreme in fact than the position it would occupy if it consisted of two like-minded individuals and one of a different orientation.

There is no direct confirmation of this general proposition. But some support comes from an intriguing study of judicial behavior on the D.C. Circuit.[112] Under Chevron v. NRDC,[113] courts are supposed to uphold agency interpretations of law so long as the interpretations are "reasonable." When do courts obey this stricture? The study strongly suggests that group polarization plays a role. The most important finding is a dramatic difference, on the United States court of appeals for the District of Columbia Circuit, between politically diverse panels (with judges appointed by Presidents of more than one party) and "unified" panels (with judges appointed by Presidents of only one party). On divided panels in which a majority of the court might be expected, on broadly speaking political grounds, to be hostile to the agency, the court

---

[111] See, e.g., Lawrence Lessig and Cass R. Sunstein, The President and the Administration, 94 Colum L Rev 1 (1994).
[112] See Frank Cross and Emerson Tiller, Judicial Partisanship and Obedience to Legal Doctrine, 107 Yale LJ 2155 (1998).
[113] See US (1984)

deferred to the agency 62% of the time. But on <u>unified</u> panels in which the court might be expected to be hostile to the agency, the court upheld the agency interpretation only 33% of the time. Note that this was the only asymmetry in the data; when courts were expected to uphold the agency's decision on political controls, they did so over 70% of the time, whether unified (71% of the time) or divided (86% of the time). There is no smoking gun here, but it seems reasonable to speculate that the seemingly bizarre result – a mere 33% validation rate in cases in which the panel was unified – reflects a process of group polarization. A group of like-minded judges may well take the relatively unusual step of rejecting an agency interpretation, whereas as a divided panel, with a check on any tendency toward extreme outcomes, is more likely to take the conventional route.

4. <u>Legislatures</u>. Legislators are likely to be susceptible to group polarization, partly because of the effects of limited argument pools, perhaps above all because of social influence (and the importance of conveying a proper signal to fellow legislators and above all constituents). Imagine, for example, that a group of Republicans and a group of Democrats are thinking about how to vote on a proposed law – perhaps involving military spending, or an increase in the minimum wage, or mandatory parental leave legislation, or greater environmental protection. If Republicans are speaking mostly with Republicans, and if Democrats are speaking mostly with Democrats, we should expect a hardening of views toward the more extreme points. Undoubtedly this is part (certainly not all) of the explanation of party-line voting. And it is easy to imagine similar effects on Congress as a whole.

A result of this general kind has been documented with the original passage of the Clean Air Act.[114] In the relevant period, there was a great deal of electoral pressure to enact some kind of clean air legislation. Both President Nixon and Senator Muskie attempted to signal to voters that they cared a great deal about the environment. The difficulty was that both of them found themselves in a kind of "politicians' dilemma," in which they had to urge more and more aggressive regulation – more aggressive, in fact, than either of them sought – precisely in order to maintain the preferred relative position vis-a-vis the electorate. Congress itself polarized accordingly, toward a more extreme position than most or even all individuals would have sought beforehand.

There are significant differences between the legislative process and the contexts in which group polarization has been studied, above all because members of Congress are subject to external political sanctions. Even if members are persuaded that a certain course of action makes best sense, they may vote otherwise, simply because of what their constituents want. Hence a limited argument pool, for members of a particular party, may matter much less than a clear signal from people back home. This point may explain occasional defections on both sides; certainly it explains why some members are able to resist both party pressures and the logic of group polarization. Unambiguous electoral signals can be a powerful buffer against that logic (though the signals themselves may be a function of group polarization within the electorate).

The same point bears on the relevance of social influence. Members of the Republican Party are likely to care a great deal what fellow Republicans think of them; but they probably

---

[114] See E. Donald Elliott et al., supra note.

care still more about what local voters think of them. To be sure, the two are not independent of one another. If a certain Republican seems like an outlier among Republicans generally – for example, if he seems less sympathetic to the religious right than his colleagues – his electoral prospects might be damaged simply by virtue of this signal. But analytically, the two are different. Here too the votes of constituents may matter more than group deliberations (taking members of the same party as the relevant group).

It is important to underline here the fact that the mechanisms of group polarization may sometimes be at work with constituents as well. We can imagine a society in which Republicans speak mostly with each other; we can imagine a society in which Democrats speak mostly with one another too. If this is the situation, polarization should occur within political camps. We might think that group polarization supplies one of the many factors behind the sharp split between Republicans and Democrats on the impeachment of President Clinton.[115]

VI. Deliberative Trouble?

A. Doubts

The phenomenon of group polarization, alongside the phenomenon of social cascades, raises severe doubts about the value of deliberation. Note here that deliberation might be justified, as a social practice, on one of two grounds. It may be that on the question at issue, there is a truth of the matter – a correct answer – and deliberation might be justified as the best way of reaching it.[116] Group decisions are more likely to be right than decisions made by individuals. Alternatively, we might favor deliberation for the opposite reason; doubting whether there is a truth of the matter, a society might seek a deliberative process on the theory that this is the only reasonable and fair way to reach a decision that will be imposed on the group.[117] Group polarization raises no difficulty for the second sort of account; but it poses real problems for the first. If the effect of deliberation is to move people toward a more extreme point in the direction of their original tendency, why is it anything to celebrate? Nor do the mechanisms provide much reason for confidence. If people are shifting their position in order to maintain their reputation and self-conception, is there any reason to think that deliberation is making things better rather than worse? If shifts are occurring as a result of partial and frequently skewed argument pools, the results of deliberative judgments may be far worse than the results of simply taken the median of predeliberation judgments.

To be sure, those who emphasize the ideals associated with deliberative democracy tend to emphasize its preconditions, which include political equality and the goal of "reaching understanding."[118] In real-world deliberations, behavior is often strategic, and equality is often

---

[115] For more detailed discussion, see Cass R. Sunstein, Group Dynamics, forthcoming in Aftermath (2000).

[116] This is the tendency in Gutmann and Thompson, supra note

[117] See the discussion of imperfect procedural justice and pure procedural justice in John Rawls, A Theory of Justice (1971).

[118] See Jurgen Habermas, A Theory of Communicative Action 99 (1984). Thus Habermas distinguishes between strategic and communicative action and stresses "the cooperatively pursued goal of reaching understanding"; compare the treatment in Gutmann and Thompson, supra note, at 52-94, emphasizing the idea of reciprocity, which emphasizes the desire to justify one's position by reference to reasons.

absent in one or another form. But group polarization is likely to occur even in the face of equality and entirely conscientious efforts at reaching both truth and understanding. The existence of a limited argument pool, strengthening the existing tendency within the group, will operate in favor of group polarization even if no individual behaves strategically. By itself this will produce group polarization whether or not social influence is operating. In any case social influences need not be inconsistent with the effort to produce truth and understanding; when people attempt to position themselves in a way that fits with their best self-conception, or their preferred self-presentation, nothing has gone wrong, even from the standpoint of deliberation's most enthusiastic defenders.[119] Perhaps group polarization could be reduced or even eliminated if we emphasized that good deliberation has full information as a precondition; but that requirement is extremely stringent, and if there is already full information, the point of deliberation is greatly reduced.[120] In any case the group polarization phenomenon suggests that in real-world situations, deliberation is hardly guaranteed to increase the likelihood of arriving at truth.

Of course we cannot say, from the mere fact of polarization, that there has been a movement in the wrong direction. Perhaps the more extreme tendency is better; recall that group polarization is likely to have fueled the antislavery movement and many others that deserve to meet with widespread approval. Extremism need not be a word of opprobrium, and in any case a group of moderates is likely, as noted, to become entrenched in its moderation by virtue of the mechanisms discussed here. In addition, group polarization can be explained partly by reference to the fact that people who are confident are likely to be persuasive; and it seems sensible to say that as a statistical matter, people who are confident are more likely to be right. But when group discussion tends to lead people to more strongly held versions of the same view with which they began, and if social influences and limited argument pools are responsible, there is little reason for great confidence in the effects of deliberation.

B. A Lesson

As a thought experiment, imagine a deliberating body consisting of all citizens in the relevant group; this may mean all citizens in a community, a state, a nation, or the world. By hypothesis, the argument pool would be very large; it would be limited only to the extent that the set of citizen views was also limited. Social influences would undoubtedly remain; hence people might shift because of a desire to maintain their reputation and self-conception, by standing in a certain relation to the rest of the group. But to the extent that deliberation revealed to people that their private position was different, in relation to the group, from what they thought it was, any shift would be in response to an accurate understanding of all relevant citizens, and not a product of a skewed group sample.

This thought experiment does not suggest that the hypothesized deliberating body would be ideal. Perhaps all citizens, presenting all individual views, would offer a skewed picture from the normative point of view. Perhaps weak arguments would be made and repeated and repeated again, while good arguments would be offered infrequently. But at least a deliberating body of

---

[119] See Robert Goodin, Laundering Preferences, in Foundations of Social Choice Theory 75, 77-90 (Jon Elster and Aanund Hyllund eds. 1986).
[120] Not eliminated. There remains the question of what to do, given a certain understanding of the facts.

all citizens would remove some of the distortions in the group polarization experiments, where generally like-minded people, not exposed to others, shift in large part because of that limited exposure.

A possible conclusion would return to the need for full information, not only about facts but also about possible options and values, and suggest that in any deliberating body, it is important to ensure a wide mix of views, so as to ensure that a distorted argument pool does not produce unearned or unjustified shift. For a leader of any institution, it makes sense, in any ordinary circumstance, to try to ensure a broad array of views, simply in order to ensure against the predictable entrenchment of private judgments. The idea of a "public sphere" can be understood as an effort to ensure a domain in which multiple views can be heard, and can be heard by people having multiple perspectives.[121] Thus there is reason for caution about any institutional practice that insulates people from competing arguments.[122] Indeed, an understanding of group polarization suggests that it would be desirable to take steps to reduce the likelihood that panels on federal courts of appeals do not consist solely of appointees of presidents of any single political party.

Of course any argument pool will be limited; no one has time to listen to every point of view. But perhaps the largest lesson involves the need for caution about the effects of deliberation within groups all or most of whose members already have an extreme tendency. Heterogeneous groups are a far better source of good judgments. The principal qualification here is that heterogeneity is by itself neither here nor there; the question is how to ensure appropriate heterogeneity. For example, it would not make sense to say that in a deliberating group attempting to think through issues of affirmative action, it is important to allow exposure to the view that slavery was good and should be restored. The constraints of time and attention call for limits to heterogeneity; and – a separate point -- for good deliberation to take place, some views are properly placed off the table, simply because they are so invidious and implausible. This point might seem to create a conundrum: To know what points of view should be represented in any group deliberation, it is important to have a good sense of the substantive issues involved, indeed a sufficiently good sense as to generate judgments about what points of view must be included and excluded. But if we already know that, why should we not proceed directly to the merits? If we already know that, before deliberation occurs, does deliberation have any point at all?

The answer is that we often do know enough to know which views count as reasonable, without knowing which view counts as right, and this point is sufficient to allow people to construct deliberative processes that should correct for the most serious problems potentially created by group deliberation. What is necessary is not to allow every view to be heard, but to ensue that no single view is so widely heard, and reinforced, that people are unable to engage in critical evaluation of the reasonable competitors. In this way an understanding of group polarization provides no simple view of deliberation as a social process, but does provide an important lesson to those interested in the construction of public spaces or a well-functioning public sphere.

---

[121] See Jurgen Habermas, The Structural Transformation of the Public Sphere 231-50 (1991).
[122] See Lessig, supra note, at 186.

Of course the provision of diverse views does not guarantee good deliberation. Among other things, most people are subject to "confirmatory bias," in accordance with which exposure to competing position will not dislodge and may even strengthen the antecedently held position.[123] On questions of morality and fairness, and undoubtedly less as well, those who listen to diverse opinions may well emerge from the experience with an enhanced belief in the soundness of their original commitment.[124] But this is not a universal phenomenon, and at least an understanding of competing views is likely to weaken the forms of fragmentation and misunderstanding that come from deliberation among the like-minded.

C. The Deliberative Opinion Poll: A Contrast

In an interesting combination of theoretical and empirical work, James Fishkin has pioneered the idea of a "deliberative opinion poll," in which small groups, consisting of highly diverse individuals, are asked to come together and to deliberate about various issues. Fishkin finds some noteworthy shifts in individual views; but he does not find a systematic tendency toward polarization. In England, for example, deliberation led to reduced interest in using imprisonment as a tool for combating crime.[125] The percentage believing that "sending more offenders to prison" is an effective way to prevent crime went down from 57% to 38%; the percentage believing that fewer people should be sent to prison increased from 29% to 44%; belief in the effectiveness of "stiffer sentences" was reduced from 78% to 65%.[126] Similar shifts were shown in the direction of greater enthusiasm for procedural rights of defendants and increased willingness to explore alternatives to prison. These are not the changes that would be predicted by group polarization. The probable reason is that in Fishkin's studies, participants were presented with a set of written materials that attempted to be balanced but that would likely move people in different directions from those that would be expected by simple group discussion. Indeed, the very effort to produce balance should be expected to shift large majorities into small ones, pressing both sides closer to 50% representation; and this is in fact what was observed.[127]

In other experiments with the deliberative opinion poll, shifts included a mixture of findings, with larger percentages of individuals concluding that legal pressures should be increased on fathers for child support (from 70% to 85%) and that welfare and health care should be turned over to the states (from 56% to 66%).[128] Indeed, on many particular issues, the effect of deliberation was to create an increase in the intensity with which people held their preexisting convictions.[129] These findings are consistent with the prediction of group polarization. But this was not a uniform pattern, and on some questions deliberation increased the percentage of people holding a minority position (with, for example, a jump from 36% to 57% of people favoring

---

[123] See Jonathan Baron, Thinking and Deciding (1994).

[124] See id.

[125] Fishkin, supra note, at 206-07.

[126] Id.

[127] See id.

[128] Fishkin and Luskin, supra note, at 23.

[129] See id. at 22-23 (showing a jump, on a scale of 1 to 4, from 3.51 to 3.58 in intensity of commitment to reducing the deficit); a jump, on a scale of 1 to 3, from 2.71 to 2.85 in intensity of support for greater spending on education; showing a jump, on a scale of 1 to 3, from 1.95 to 2.16, in commitment to aiding American business interests abroad).

policies making divorce "harder to get").[130] Taken as a whole, a great deal of Fishkin's data seem to support the group polarization hypothesis; what does not is probably a product of some combination of statistical noise, effects of external presentation, and deviations produced by members of the particular groups involved.

Conclusion

In this essay I have attempted to discuss the phenomenon of group polarization and to trace some of its implications for law and political theory. The basic point is that group deliberation can create polarization of both groups and individuals. The underlying mechanisms have a great deal to do with skewed and limited argument pools, and with people's desire to maintain relative position of a certain kind (perhaps as a heuristic, perhaps for reputational reasons, perhaps because of self-conception). Group polarization can occur on juries, within legislatures, and on multimember courts and commissions. The phenomenon helps explain why many groups go, quite surprisingly, in extreme directions.

In the abstract, and without knowing about the underlying substance, it is impossible to say whether this tendency is good or bad. But the mechanisms that underlie group polarization raise serious questions about any general enthusiasm for deliberative processes. If the argument for deliberation is that it is likely to yield correct answers to social questions, group polarization suggests the need for attention to the background conditions in which this is likely to be the case. Like-minded people, engaged in discussion with one another, may lead each other in the direction of error and falsehood, simply because of the limited argument pool and the operation of social influences. I have suggested that the best response to this problem is to attempt to ensure against social balkanization and fragmentation, through mechanisms providing a "public sphere" that is used, at once, by people with competing perspectives on facts and values. If a general public sphere is unavailable or not feasible, it becomes all the more important to ensure that in the course of deliberation, people are exposed to a range of reasonable competing views.

Of course it might seem hard to know what counts as a reasonable competing view without knowing what is actually right, and if we already know that, there might seem to be little point to deliberation. But short of knowing what is right, it is possible to know something about the range of reasonable candidates, and about who might learn from whom. Perhaps the largest lesson provided by group polarization involves the need to structure processes of deliberation so as to ensure that people are exposed, not to softer or louder echoes of their own voices, but to a range of reasonable alternatives. By itself, that lesson is very far from new; but an understanding of the potential effects of group polarization argues in favor of fresh thinking, and possible reforms, in many contemporary institutions.

---

[130] Id. at 23. See also id at 22 (showing an increase , on a scale of 1 to 3, from 1.40 to 1.59 in commitment to spending on foreign aid ; also a decrease, on a scale of 1 to 3, from 2.38 to 2.27 in commitment to spending on social security).

Readers with comments should address them to:

Cass R. Sunstein
Karl N. Llewellyn Distinguished Service Professor
University of Chicago Law School
1111 East 60th Street
Chicago, IL 60637

Chicago Working Papers in Law and Economics
(Second Series)

1. William M. Landes, Copyright Protection of Letters, Diaries and Other Unpublished Works: An Economic Approach (July 1991).
2. Richard A. Epstein, The Path to *The T. J. Hooper*: The Theory and History of Custom in the Law of Tort (August 1991).
3. Cass R. Sunstein, On Property and Constitutionalism (September 1991).
4. Richard A. Posner, Blackmail, Privacy, and Freedom of Contract (February 1992).
5. Randal C. Picker, Security Interests, Misbehavior, and Common Pools (February 1992).
6. Tomas J. Philipson & Richard A. Posner, Optimal Regulation of AIDS (April 1992).
7. Douglas G. Baird, Revisiting Auctions in Chapter 11 (April 1992).
8. William M. Landes, Sequential versus Unitary Trials: An Economic Analysis (July 1992).
9. William M. Landes & Richard A. Posner, The Influence of Economics on Law: A Quantitative Study (August 1992).
10. Alan O. Sykes, The Welfare Economics of Immigration Law: A Theoretical Survey With An Analysis of U.S. Policy (September 1992).
11. Douglas G. Baird, 1992 Katz Lecture: Reconstructing Contracts (November 1992).
12. Gary S. Becker, The Economic Way of Looking at Life (January 1993).
13. J. Mark Ramseyer, Credibly Committing to Efficiency Wages: Cotton Spinning Cartels in Imperial Japan (March 1993).
14. Cass R. Sunstein, Endogenous Preferences, Environmental Law (April 1993).
15. Richard A. Posner, What Do Judges and Justices Maximize? (The Same Thing Everyone Else Does) (April 1993).
16. Lucian Arye Bebchuk and Randal C. Picker, Bankruptcy Rules, Managerial Entrenchment, and Firm-Specific Human Capital (August 1993).

36. Cass R. Sunstein, Social Norms and Social Roles (January 1996).
37. J. Mark Ramseyer and Eric B. Rasmusen, Judicial Independence in Civil Law Regimes: Econometrics from Japan (January 1996).
38. Richard A. Epstein, Transaction Costs and Property Rights: Or Do Good Fences Make Good Neighbors? (March 1996).
39. Cass R. Sunstein, The Cost-Benefit State (May 1996).
40. William M. Landes and Richard A. Posner, The Economics of Legal Disputes Over the Ownership of Works of Art and Other Collectibles (July 1996).
41. John R. Lott, Jr. and David B. Mustard, Crime, Deterrence, and Right-to-Carry Concealed Handguns (August 1996).
42. Cass R. Sunstein, Health-Health Tradeoffs (September 1996).
43. Douglas G. Baird, The Hidden Virtues of Chapter 11: An Overview of the la and Economics of Financially Distressed Firms (March 1997).
44. Richard A. Posner, Community, Wealth, and Equality (March 1997).
45. William M. Landes, The Art of Law and Economics: An Autobiographical Essay (March 1997).
46. Cass R. Sunstein, Behavioral Analysis of Law (April 1997).
47. John R. Lott, Jr. and Kermit Daniel, Term Limits and Electoral Competitiveness: Evidence from California's State Legislative Races (May 1997).
48. Randal C. Picker, Simple Games in a Complex World: A Generative Approach to the Adoption of Norms (June 1997).
49. Richard A. Epstein, Contracts Small and Contracts Large: Contract Law through the Lens of Laissez-Faire (August 1997).
50. Cass R. Sunstein, Daniel Kahneman, and David Schkade, Assessing Punitive Damages (with Notes on Cognition and Valuation in Law) (December 1997).
51. William M. Landes, Lawrence Lessig, and Michael E. Solimine, Judicial Influence: A Citation Analysis of Federal Courts of Appeals Judges (January 1998).
52. John R. Lott, Jr., A Simple Explanation for Why Campaign Expenditures are Increasing: The Government is Getting Bigger (February 1998).

70. Cass R. Sunstein, Must Formalism Be Defended Empirically? (March 1999)

71. Jonathan M. Karpoff, John R. Lott, Jr., and Graeme Rankine, Environmental Violations, Legal Penalties, and Reputation Costs (March 1999)

72. Matthew D. Adler and Eric A. Posner, Rethinking Cost-Benefit Analysis (April 1999)

73. John R. Lott, Jr. and William M. Landes, Multiple Victim Public Shooting, Bombings, and Right-to-Carry Concealed Handgun Laws: Contrasting Private and Public Law Enforcement (April 1999)

74. Lisa Bernstein, The Questionable Empirical Basis of Article 2's Incorporation Strategy: A Preliminary Study (May 1999)

75. Richard A. Epstein, Deconstructing Privacy: and Putting It Back Together Again (May 1999)

76. William M. Landes, Winning the Art Lottery: The Economic Returns to the Ganz Collection (May 1999)

77. Cass R. Sunstein, David Schkade, and Daniel Kahneman, Do People Want Optimal Deterrence? (June 1999)

78. Tomas J. Philipson and Richard A. Posner, The Long-Run Growth in Obesity as a Function of Technological Change (June 1999)

79. David A. Weisbach, Ironing Out the Flat Tax (August 1999)

80. Eric A. Posner, A Theory of Contract Law under Conditions of Radical Judicial Error (August 1999)

81. David Schkade, Cass R. Sunstein, and Daniel Kahneman, Are Juries Less Erratic than Individuals? Deliberation, Polarization, and Punitive Damages (September 1999)

82. Cass R. Sunstein, Nondelegation Canons (September 1999)

83. Richard A. Posner, The Theory and Practice of Citations Analysis, with Special Reference to Law and Economics (September 1999)

84. Randal C. Picker, Regulating Network Industries: A Look at *Intel* (October 1999)

85. Cass R. Sunstein, Cognition and Cost-Benefit Analysis (October 1999)

Group Polarization and *12 Angry Men*

Cass R. Sunstein

Deliberating groups, including juries, typically end up in a more extreme position in line with their predeliberation tendencies. A jury whose members are inclined, before deliberation, to find a defendant not guilty will likely render a verdict of not guilty; a jury whose members want to award punitive damages will likely produce an award higher than that of the median juror. The phenomenon of group polarization, found in many domains, stems from a combination of information pooling and peer pressure. The events portrayed in the film 12 Angry Men *seem to defy the logic of group polarization, but the film nonetheless shows an acute psychological sense.*

Key words: negotiation, group polarization, peer pressure, deliberation, juries, film.

Introduction

One of the most striking features of the film *12 Angry Men* is its apparent inconsistency with what social scientists know about group dynamics. True, Henry Fonda is an impressive guy, in this movie as in every other. But a single juror is highly unlikely to be able to move eleven people who are leaning the other way. In fact, the usual pattern is just the opposite: *If almost all members of a twelve-member jury are inclined to convict the defendant, the jury is overwhelmingly likely to convict that defendant.*

Cass R. Sunstein is the Karl N. Llewellyn Distinguished Service Professor at the University of Chicago Law School and the department of political science at the University of Chicago. His e-mail address is csunstei@uchicago.edu.

My goal here is to examine why this is so — and to connect that examination with an inquiry into the social dynamics of the jury in *12 Angry Men*. It turns out that the movie is not quite as unrealistic as it seems (Brown 1985). To appreciate that conclusion, let us begin with some political issues.

Polarizing Groups

In the summer of 2005, my colleagues and I conducted a small experiment in democracy in Colorado (Hastie, Schkade, and Sunstein Forthcoming). About sixty American citizens were brought together and assembled into ten "juries," mostly consisting of six people. Members of each "jury" were asked to deliberate on three of the most controversial issues of the day:

- Should states allow same-sex couples to enter into civil unions?

- Should employers engage in "affirmative action" by giving a preference to members of traditionally disadvantaged groups?

- Should the United States sign an international treaty to combat global warming?

As the experiment was designed, the groups consisted of "liberals" from Boulder and "conservatives" from Colorado Springs. In election-year parlance, they comprised five "Blue State" groups, whose members initially tended toward liberal positions in general, and five "Red State" groups, whose members tended generally toward conservative positions. People were asked to state their opinions anonymously both before and after fifteen minutes of group discussion.

What was the effect of discussion? In almost every jury, people were moved toward a more extreme position in line with the original tendency within the group. "Outliers" — those whose opinions, like Juror 8 portrayed by Henry Fonda in *12 Angry Men*, diverged significantly from the majority — were never to able to make much progress; they were more likely to become converts than converters. More specifically, discussion made civil unions even more popular among liberals and even less popular among conservatives. Liberals favored an international treaty to control global warming before discussion; they favored it more strongly after discussion. Conservatives were slightly negative about that treaty before discussion; they strongly opposed it after discussion. Mildly favorable toward affirmative action before discussion, liberals became strongly favorable toward affirmative action after discussion. Firmly negative about affirmative action before discussion, conservatives became even more negative about affirmative action after discussion.

The Colorado experiment is an illustration of the well-established phenomenon of group polarization, by which members of deliberating groups typically end up in a more extreme position in line with their

tendencies before deliberation began (Brown 1985). Within small groups, those with unusual or contrary positions are rarely able to move other group members in their preferred direction. In fact, group polarization has been found in hundreds of studies involving people in more than a dozen countries (Brown 1985). In one study, white people who were not inclined to show racial prejudice showed *less* prejudice after deliberation than before; but white people who were inclined to show such prejudice displayed even *more* prejudice after deliberation (Myers and Bishop 1970). Similarly, after deliberation, French people became more distrustful of the United States and its intentions with respect to foreign aid (Brown 1985). In the same vein, feminism became more attractive to women after internal discussions, at least if those women were antecedently inclined to favor feminism (Myers 1975).

In the domain of law, as well, ample evidence attests to the pervasiveness of group polarization. Group polarization can occur for judgments of guilt and for sentencing in criminal cases (Myers and Kaplan 1976). Studies have shown that when a strong majority of jurors are inclined to convict, the jury is likely to convict. Across the general run of cases, there is essentially zero probability that a group of eleven to one will shift to zero to twelve (Brown 1985) — which makes accepting the verisimilitude of the events depicted in *12 Angry Men* somewhat problematic.

In punitive damage cases as well, deliberating juries have been found to polarize, producing awards that are often higher than those of the median juror before deliberation begins (Schkade, Sunstein, and Kahneman 2000). When individual jurors begin with a high degree of moral outrage about a defendant's conduct, juries become more outraged after deliberation than their median member had been before deliberation; dollar awards are often as high as or even higher than the highest award favored, before deliberation, by any individual juror. For example, Schkade, Sunstein, and Kahneman (2000) found in 27 percent of the cases they studied that the jury's award was as high as or higher than that favored by any individual juror before deliberation.

Why Polarization?

Why does group polarization occur? There are several explanations (see Brown 1985). The first and most important involves the exchange of information. Fortunately, most people *do* listen to the arguments made by other people. In any group whose members have a definite initial inclination, the views of most people in the group will (by definition) be skewed in the direction of that inclination. Suppose that the majority position within a group is that a defendant should be convicted, that global warming is a serious problem, or that the incumbent president is doing a terrific job. As a statistical matter, the arguments favoring that initial position will be more numerous than the arguments pointing in the other direction. Individuals

will have heard of some, but not all, of the arguments that emerge from group deliberation. As a result of hearing the various arguments, deliberation will lead people toward a more extreme point in line with what group members initially believed. Through this process, many minds can polarize, and in exactly the same direction.

The second explanation involves social comparison. People usually want to be perceived favorably by other group members, even on a jury. Sometimes people's publicly stated views are, to a greater or lesser extent, a function of how they want to present themselves and to be perceived. Once they hear what others believe, some will adjust their positions at least slightly in the direction of the dominant position. In a left-wing group, for example, those who lean to the left will be more acceptable, and for this reason they might well end up leaning somewhat more to the left. So too on a jury: most people do not want to be perceived as silly or stupid, so if eleven people are inclined to convict a defendant, the twelfth will usually go along.

The third explanation of group polarization stresses the close links among confidence, extremism, and corroboration by others (Baron et al. 1996). As people gain confidence, they usually become more extreme in their beliefs. Agreement from others tends to increase confidence, and for this reason like-minded people, having deliberated with one another, become more extreme as they become less tentative. In many contexts, people's opinions become more extreme simply because their views have been corroborated, and because they become more confident after learning that others share their views (Baron et al. 1996). So, too, on a criminal jury: if ten people want to convict a defendant, and two others are unsure, their proconviction inclinations will be strengthened after corrobation. If the two others are Juror 8, they must have a lot of confidence to resist the impact of the other jurors' views.

For juries, the central lesson is clear. If most jury members believe that a defendant is guilty, the jury is likely to convict that defendant, simply because it will polarize in the direction of a guilty verdict. The principal qualification involves the "beyond a reasonable doubt" standard; that standard creates a deliberate skew in the defendant's favor, and that skew will also affect the jury (MacCoun and Kerr 1988). But the skew is most unlikely to save the defendant if only one or two jurors believe that he or she is innocent.

Realism and *12 Angry Men*

At this point, one might easily conclude that *12 Angry Men* is a bit of a fairy tale — gripping to watch, and entertaining too, but psychologically implausible. On the other hand, simple intuition and casual empiricism suggest that it *is* possible for one person to sway eleven others. As one example, consider the tale of the *Emperor's New Clothes*, in which a small boy, having said what he sees, is able to turn around a large group of people who have operated under (what they know in their hearts to be) a fiction. But

Juror 8's task is much harder. He must convince people who actually believe that the defendant is guilty.

To understand why he succeeds, and why the movie works, notice three simple points. First, the most guilt-inclined jurors have very little information, and Juror 8 has much more, or at least he is able to ask the right questions. In most groups, the distribution of initial information strongly favors the position that most people initially hold; this is not true in the context of *12 Angry Men*. Second, Juror 8 is a firm, highly likable, rational, and generally confident dissenter, relatively impervious to the ordinary kinds of social influence that dampen minority positions (Sunstein 2003). Third, the least competent jurors are the most certain that the defendant is guilty; the most competent jurors are the most ambivalent (Brown 1985). The audience can plainly see this fact and hence the stage is set for a reversal of opinion: if Juror 8 can convince the most competent jurors, who are already ambivalent, he might be able to create a cascade in the defendant's favor.

If we want to know how most groups, including most juries, tend to work, we will not take *12 Angry Men* very seriously. Group polarization is the human regularity, and deliberation rarely leads group members to abandon their initial inclination. But however rare, exceptions are important and revealing. *12 Angry Men* works in large part not because the outcome is realistic, but because the movie shows a realistic and even acute understanding of the kinds of processes that produce the rare exceptions.

Those rare exceptions are exceedingly important. Over the long arc of history, those with a minority position have often proved able to convince the majority; consider the movements for abolishing slavery, civil rights, women's suffrage, and environmental protection. Group polarization is the general rule, but *12 Angry Men* has the signal virtue of casting a clear light on the social processes by which one or a few can indeed convince the many.
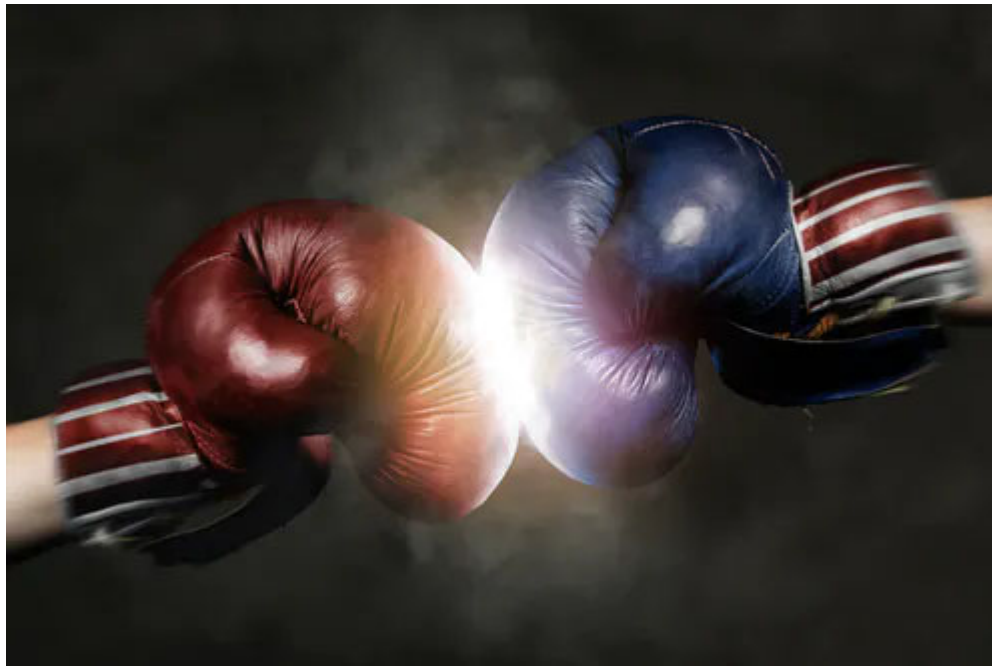
REFERENCES

Baron, R., S. Hoppe, B. Linneweh, and D. Rogers. 1996. Social corroboration and opinion extremity. *Journal of Experimental Social Psychology* 32: 537–560.
Brown, R. 1985. *Social psychology: The second edition*. New York: The Free Press.
Hastie, R., D. Schkade, and C. R. Sunstein. Forthcoming. What really happened on deliberation day? *California Law Review* 95.
MacCoun, R. J. and N. Kerr. 1988. Asymmetric influences in mock jury deliberation: Jurors' bias for leniency. *Journal of Personality and Social Psychology* 54(1): 21–33.
Myers, D. G. 1975. Discussion-induced attitude polarization. *Human Relations* 28(10): 699–714.
Myers, D. G. and G. D. Bishop. 1970. Discussion effects on racial attitudes. *Science* 169(3947): 778–779.
Myers, D. G. and M. F. Kaplan. 1976. Group-induced polarization in simulated juries. *Personality and Social Psychology Bulletin* 2: 63–66.
Schkade, D., C. R. Sunstein, and D. Kahneman. 2000. Deliberating about dollars: The severity shift. *Columbia Law Review* 100(4): 1139–1176.
Sunstein, C. R. 2003. *Why societies need dissent*. Cambridge, MA: Harvard University Press.

THE C◯NVERSATION

Academic rigor, journalistic flair

Animosity between partisan voters has grown in recent years. Gutzemberg/Shutterstock.com

Political polarization is about feelings, not facts

July 31, 2019 7.43am EDT

Politicians and pundits from all quarters often lament democracy's polarized condition.

Similarly, citizens frustrated with polarized politics also demand greater flexibility from the other side.

Decrying polarization has become a way of impugning adversaries. Meanwhile, the political deadlock and resentment that polarization produces goes unaddressed. Ironic, right?

Commentators rarely say what they mean by polarization. But if Americans are to figure out how to combat it, they need to begin from a clear understanding of what polarization is.

My forthcoming book, "Overdoing Democracy," argues that polarization isn't about where you get your news or how politicians are divided – it's about how a person's political identity is wrapped up with almost everything they do.
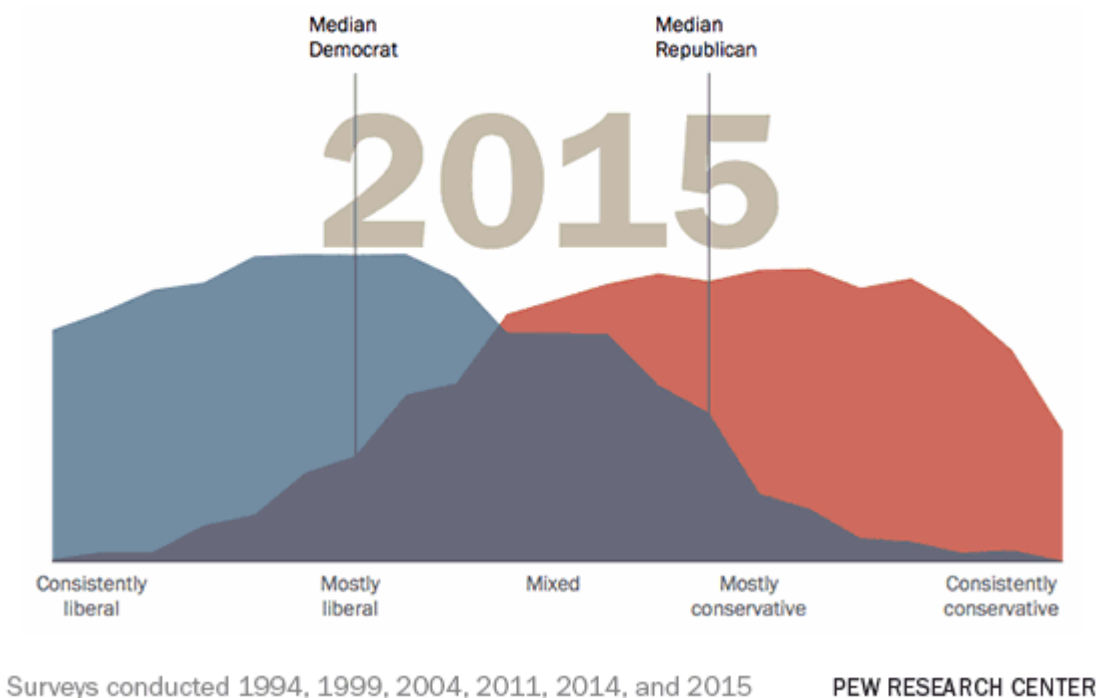
Author

Robert B. Talisse
W. Alton Jones Professor of Philosophy,
Vanderbilt University

Political polarization in the American public, 1994 - 2015

Surveys conducted 1994, 1999, 2004, 2011, 2014, and 2015   PEW RESEARCH CENTER

Polarization, three ways

Start with the obvious: Polarization is the political distance separating partisans. But this intuitive idea is not so simple, as political scientists have at least three ways of measuring political distance.

One compares the platforms of competing parties. Polarization is the extent to which these are opposed.

A second assesses each party's ideological homogeneity. This definition of polarization concerns how many of the party's officials are "moderates" or bridge-builders.

A third involves neither platforms nor officials, but instead the emotions of ordinary citizens who affiliate with a political party. It tracks the extent to which citizens dislike affiliates of other parties.

Research suggests that, although the major U.S. parties are severely polarized along the first two dimensions, the American public is no more divided now over policy than it was 30 years ago. In fact, on certain hot-button issues such as abortion and gay rights, rank-and-file citizens who identify with a political party have moved closer together.

Nonetheless, Americans believe that their policy divisions are especially pronounced. Polarization in the third sense has skyrocketed with interparty animosity more intense now than it has been for the past 25 years.

What each party thinks of the other

In a 2016 survey, many Democrats and Republicans said that members of the other party were more likely to be closed-minded, dishonest and immoral, compared to other Americans.

	Republicans say Democrats are more…	Democrats say Republicans are more…
Closed-minded	52%	70%
Dishonest	45%	42%
Immoral	47%	35%
Unintelligent	32%	33%
Lazy	46%	18%

Chart: The Conversation, CC-BY-ND • Source: Pew • Get the data

In other words, though Americans are less divided over the issues, we see ourselves as profoundly at odds. We more intensely dislike those we regard as politically different from ourselves.

This suggests to me that, when citizens detest those with opposing affiliations, political parties are driven to overstate their differences, stress ideological purity and vilify the opposition.

For example, consider the popular slur among Republicans, "RINO," – or Republican In Name Only – which derides GOP members who are seen to be insufficiently devoted to the party line.

A similar dynamic can be seen in discussions of those vying for the Democratic nomination, where hopefuls are often assessed according to the extent of their anti-Trump sentiments.

And just a few days ago, the president declared that certain Democratic House congresswomen are "dangerous" and may "hate America."

Thinking as a group

Here's an easy fix to this kind of polarization: Stop hating your political adversaries. But that's easier said than done.

Why do people despise those who are politically different from themselves?

The answer lies with a widespread cognitive phenomenon called group polarization. When you talk only to those you agree with, or listen only to news that affirms your opinions, you become more radical in your beliefs.

As people radicalize like this, they grow less able to comprehend opposing views, more likely to dismiss objections to their opinions and increasingly prone to regarding dissenters as incompetent and depraved.

Recall the last time you were present in a packed arena watching your favorite team win a home game. As you roared along with your fellow fans, everyone's enthusiasm for the team spiked. At the same time, animosity for the opposing team and its fans intensified. Your mood was elevated and your identity was affirmed. Cheering with fellow fans makes us feel good about ourselves.

Echo chambers

Online environments function as immense polarization machines. They enable individuals to select their information sources and filter out challenging or unfamiliar messages.

Many have suggested that people would become less polarized if they could only break out of their "echo chambers" and expose themselves to more diverse opinions.

Trust in media

A 2014 survey asked consistently liberal and consistently conservative Americans whether they trusted different news sources.

Overall   Consistently liberal   Consistently conservative

	0%	20%	40%	60%	80%	100%
CNN						
NBC News						
ABC News						
CBS News						
Fox News						
PBS						
MSNBC						
BBC						
New York Times						
USA TODAY						
Wall Street Journal						
NPR						
Washington Post						
Google News						
Yahoo News						
Huffington Post						
Daily Show						
Colbert Report						
New Yorker						
Economist						
Sean Hannity Show						
Rush Limbaugh Show						
Bloomberg						
Glenn Beck Program						
Al Jazeera America						
Drudge Report						
Guardian						
Politico						
Mother Jones						
TheBlaze						
Slate						
Breitbart						
Ed Schultz Show						
Daily Kos						
ThinkProgress						
BuzzFeed						

Respondents categorized as mostly liberal, mixed, or mostly conservative not shown.

Chart: The Conversation, CC-BY-ND • Source: Pew Research Center • Get the data

However, there's a crucial difference between prevention and cure. Diversifying your media diet could help to prevent group polarization, but it may not reverse the polarization once it has taken effect.

A 2018 social media study exposed both Democrats and Republicans to Twitter messages from people with moderate, but opposing, viewpoints. By the end, participants actually expressed more partisan views than they had when the study began. Once group polarization has taken effect on a person, they tend to regard the expression of opposing viewpoints as an attack on their identity, and this affirms their negative attitude toward their political opposition.

People radicalize in concert with like-minded others due to the mutual affirmation of a shared identity. This behavior intensifies their shared attitudes, including a negative view of outsiders. This, in turn, generates the polarization of party platforms and officials.

From my perspective, there's no easy fix. The trouble lies with people regarding political affiliations as group identities, and their political parties as warring teams in a winner-take-all death match.

[*Deep knowledge, daily*. Sign up for The Conversation's newsletter.]

🏷 **Psychology**   **Media**   **Emotions**   **Republicans**   **Democrats**   **Polarization**   **Filter bubble**   **Political polarization**
Echo chambers

2019

A Mediator's Ethical Responsibility in Cases Involving Sexual Harassment

CO-SPONSORED BY

EDNY ADR DEPARTMENT | COLUMBIA LAW SCHOOL
ETHICS COLLOQUIUM

UNITED STATES FEDERAL DISTRICT COURT FOR THE EASTERN DISTRICT OF NEW YORK

AGENDA

A Mediator's Ethical Responsibility in Cases Involving Sexual Harassment

THURSDAY, MARCH 28, 2019

5:00PM – 5:10PM   WELCOME AND OPENING REMARKS

5:10PM – 5:25PM   ETHICAL GUIDELINES FOR MEDIATORS

5:25PM – 5:40PM   SEXUAL HARASSMENT OVERVIEW

5:40PM – 6:55PM   MEDIATING SEXUAL HARASSMENT:
                         DECISION POINTS FOR MEDIATORS

6:55PM – 7:00PM   Q&A AND CLOSING REMARKS

TRAINER BIOGRAPHIES

PROFESSOR ALEXANDRA CARTER

Alexandra Carter is a Clinical Professor of Law and the Director of the Mediation Clinic at Columbia Law School. From 2012-2016, she also served as the Law School's Director of Clinical Programs, providing leadership in the area of experiential teaching and learning.

Professor Carter's teaching, research and publications lie in the field of alternative dispute resolution, primarily in mediation and negotiation. She has been sought as a trainer on mediation, civil procedure, negotiation and dispute systems design for many different groups from private and public sectors, including the United Nations, U.S. courts and federal agencies, private corporations and law firms. She has addressed the Chinese Academy of Social Sciences Rule of Law Conference, the 5th World Peace Conference in Jakarta, the ICU-CLS Peace Summit in Tokyo, and the Ceará Supreme Court Conference on Mediation; and has contributed as a faculty speaker at universities in South America, Asia and Europe. She has served on the Alternative Dispute Resolution Committee for the New York City Bar Association, as well as the Mediator Ethics Advisory Committee for the New York State Unified Court System.

Prior to joining the Columbia faculty, Professor Carter was associated with Cravath, Swaine & Moore LLP, where she worked as part of a team defending against a multibillion dollar securities class action lawsuit related to the Enron collapse, served as the senior antitrust associate on several multibillion dollar mergers, and handled cases involving copyright law.

Professor Carter received her Juris Doctor degree in 2003 from Columbia Law School, where she earned James Kent and Harlan Fiske Stone academic honors. She also won the Jane Marks Murphy Prize for clinical advocacy and the Lawrence S. Greenbaum Prize for the best oral argument in the 2002 Harlan Fiske Stone Moot Court Competition. After earning her degree, Professor Carter clerked for the Hon. Mark L. Wolf, U.S. District Court for the District of Massachusetts in Boston.

PROFESSOR CARTER'S RESEARCH TEAM

HEIDI L. GUZMAN

Heidi L. Guzmán is a third-year student at Columbia Law School. Aside from her participation in the Mediation Clinic, Heidi is also a member of the Latinx Law Students Association and a Notes Editor for the Columbia Journal of Race and Law.

LAUREN MATLOCK-COLANGELO

Lauren Matlock-Colangelo is a third-year student at Columbia Law School. She is a member of the Mediation Clinic and a Notes Editor for the Columbia Law Review.

SHINJI RYU

Shinji Ryu is a third-year student at Columbia Law School. Shinji is a member of the Mediation Clinic, an Online Editor for the Columbia Law Review, and a member of the Asian Pacific American Law Students Association.

ADDITIONAL READINGS

I.   MEDIATION OF SEXUAL-HARASSMENT CASES

1. Jan Frankel Schau, *Where Confidentiality and Transparency Collide: In Sexual Harassment Cases, Mediators Face a Modern-Day Dilemma*, 25 DISPUTE RESOL. MAG., Winter 2019, at 6.

2. Susan K. Hippensteele, *Mediation Ideology: Navigating Space from Myth to Reality in Sexual Harassment Dispute Resolution*, 15 J. GENDER SOC. POL'Y & L. 43 (2006).

3. Robert Lewis, *Mediation of a Sexual Harassment Claim*, 24 HOFSTRA LAB. & EMP. L.J. 87 (2006).

4. Howard Gadlin, *Careful Maneuvers: Mediating Sexual Harassment*, 7 NEGOT. J. 139 (1991).

5. Barbara J. Gazeley, *Venus, Mars, and the Law: On Mediation of Sexual Harassment Cases*, 33 WILLAMETTE L. REV. 605 (1997).

6. Susan A. FitzGibbon, *Arbitration, Mediation, and Sexual Harassment*, 5 PSYCHOL. PUB. POL'Y & L. 693 (1999).

7. Linda Stamato, *Sexual Harassment in the Workplace: Is Mediation an Appropriate Forum?*, 10 MEDIATION Q. 167 (1992).

8. Mori Irvine, *Mediation: Is It Appropriate for Sexual Harassment Grievances?*, 9 OHIO ST. J. ON DISPUTE RESOL. 27 (1993).

9. Carrie A. Bond, Note, *Shattering the Myth: Mediating Sexual Harassment Disputes in the Workplace*, 65 FORDHAM L. REV. 2489 (1997).

II.   SEXUAL HARASSMENT, TRAUMA, AND PSYCHOLOGY

1. Claudia Avina & William O'Donohue, *Sexual Harassment and PTSD: Is Sexual Harassment Diagnosable Trauma?*, 15 J. TRAUMATIC STRESS 69 (2002).

2. David Isaacs, Editorial, *Sexual Harassment*, 54 J. PAEDIATRICS & CHILD HEALTH 341 (2018).

3. Thomas A. Kernodle, *Workplace Sexual Harassment*, *in* 4 SAGE ENCYCLOPEDIA OF PSYCHOLOGY AND GENDER 1852 (Kevin L. Nadal ed. 2017).

DISPUTE RESOLUTION
MAGAZINE

#MeToo

**Keeping things confidential
in sexual harassment cases**

**Non-disclosure agreements:
What do we do now?**

**Realizing restoration
for victims and offenders**

Published by the American Bar Association Section of Dispute Resolution

DISPUTE RESOLUTION
MAGAZINE

Volume 25, Number 2
WINTER 2019

Features

Articles

Digital and online editions of *Dispute Resolution Magazine* are available to members of the ABA Section of Dispute Resolution at www.americanbar.org/dispute. Join the ABA Section of Dispute Resolution to receive all the benefits of membership, including quarterly issues of *Dispute Resolution Magazine*.

Where Confidentiality and Transparency Collide

In sexual harassment cases, mediators face a modern-day dilemma

By Jan Frankel Schau

We have now gone far beyond the "he said, she said" controversy in sexual harassment cases. Reports of complaints about Hollywood producer Harvey Weinstein first appeared in the *New York Times* in early October 2017, prompting hundreds of women to come forward with other accounts of unlawful and unwanted sexual misconduct by bosses and coworkers and teachers and coaches in the entertainment industry, workplaces, schools, government offices, and many other venues. Both those who were harmed and the public have demanded full investigations and complete accountability, even when the alleged misconduct took place decades ago. In this #MeToo era, a simple denial of misconduct or a claim that the behavior was consensual is no longer enough to make allegations go away.

Today victims of sexual harassment and misconduct are filing lawsuits, demanding compensation, and publicly sharing details of the abuses. These actions often provide comfort for others who have experienced similar treatment and help create safer workplaces. As the number of allegations swells, mediators are increasingly being called upon to help parties settle these disputes before lawsuits go to trial. This may be encouraging for mediators and for the field, but it raises new practice questions and causes tension for mediators who must grapple with employers' (and often their employees') desire to settle these claims confidentially and the public's demand for transparency and accountability.

Some of the new key questions

One of the basic principles of mediation is that the participants should feel free to communicate intimate, personal, and yet-unproven details that underlie their claims. Under many state statutes and court rules, anything said or any admission made for the purpose of, in the course of, or pursuant to a mediation is protected from discovery and admissibility as evidence. For example, the California Evidence Code prohibits the mediator from testifying in any subsequent civil proceeding.[1] (For a discussion about a revision of California's confidentiality rule, see "On Professional Practice" on page 30 in this issue.)

As mediators, knowing that we will never be called upon to take sides in a dispute that came before us or accurately recall all that was said during a mediation (much of which is not memorialized in writing) gives us considerable comfort.

But what happens to that comfort if we are called upon to mediate a series of disputes against the same employer, prompted by multiple claimants who allege that the company's powerful leaders have repeatedly taken sexual advantage of the young women who work for them? When many allegations are repeated against the same bosses, is the duty of confidentiality stronger than an individual's personal sense of moral duty to disclose ongoing harm? It is noteworthy that, as a general principle in cases involving allegations about workplace misconduct, the individual alleged perpetrator is seldom present at the mediation, but

the company and its lawyers (and often times insurance carriers) are there to do "damage control" and settle these matters before they become public.

If the parties expressly agree that their settlement is not confidential, does the mediator then have the responsibility — or even the legal competence — to testify about the negotiations leading up to the settlement or the ultimate settlement terms? Are the statutes that so carefully protect the confidentiality of the mediation process now overridden by the interests of the public or other potential victims?

The purpose of offering confidentiality is to provide a safe place in which disputants may air their grievances and resolve their differences without a judge, a jury, or the public listening in. Is this fair, just, and right in the context of sexual harassment cases? Are mediators playing a role in obfuscating the public's "right to know" about a history of sexual assault or mistreatment? Are we making a politically-based judgment that acts of sexual harassment must be disclosed, whereas discrimination based on race or disability, for example, is entitled to the full protections of confidentiality? It is an ethical conundrum — but one that has solutions for the creative mediator.

Confidentiality: one of mediation's basic principles

In employment disputes, both sides typically agree — or at least accede to the employer's demand — to include a confidentiality clause in the settlement agreement. (This is in addition to the mediator's customary Agreement to Participate, which usually includes language explaining that the process is confidential.) The confidentiality clause in the settlement agreement will usually identify a few necessary exclusions under the law (such as enforcement of the agreement, information to attorneys and accountants, or disclosure where required by other lawful process). Occasionally, the parties may negotiate a specific statement to be used in the event that either party is asked to comment on the settlement publicly. (For a more extended discussion of nondisclosure agreements, see the article by Elizabeth Tippett on page 12.)

While those who favor prohibiting non-disclosure clauses in sexual harassment settlements see a strong value in exposing abusers so that other potential

> There is good reason to keep the salacious details involved in many sexual harassment cases out of the public eye. Yet by doing so, the process of mediation can become a legal way to 'hush' the victims …

victims (as well as those who have been wronged by the same "bad actor") will be aware of the claims, from both the victim and perpetrator's point of view there is a competing value in keeping the claims private. Many parties and their lawyers choose mediation as a means of resolving their disputes expressly because they want to avoid the publicity, anxiety, risk, and expense of discussing the disputes in open court. By their nature, these types of cases are deeply personal. They may include the most intimate of details and may expose those involved to personal and professional backlash that may extend far beyond the payment of damages.

As we know from news reports, these kinds of claims can be deeply embarrassing, bringing shame and hurting the wrongdoer's reputation as well as damaging marriages and other relationships. They can result in loss of employment and, in the case of Harvey Weinstein, may even result in criminal charges.

There is good reason to keep the salacious details involved in many sexual harassment cases out of the public eye. Yet by doing so, the process of mediation can become a legal way to "hush" the victims, to keep them from disclosing the facts or outcome to potential claimants, other victims, or even just citizens who may be interested in knowing about this conduct by, for example, their elected officials, clergy, or other public figures.

Employers and employees have effectively used confidentiality clauses as part of the settlement of employment-related disputes arising out of allegations of sexual harassment. Yet the combination of a new tax law and the societal pressure to expose the misconduct may confound and obfuscate that process.

A new consideration: 'the Harvey Weinstein tax'

Until the end of 2017, federal tax law allowed businesses and employees to claim a business expense deduction for litigation arising out of claims of sexual harassment if the settlement of those claims was kept confidential. But federal tax legislation filed in 2017, known as "the Harvey Weinstein tax" and prompted by the political movement toward disclosure and the wish to provide information for future claimants, disallowed this deduction. Any payment on account of alleged sexual harassment or abuse may not be deducted as a business expense if it is subject to a confidentiality provision. If the parties expressly hold that their agreement is confidential, the employer cannot deduct the payment as an ordinary business expense and the claimant may not deduct her attorney's fees as an expense.

In practical terms, confidentiality is often a huge incentive, providing leverage for the plaintiff and her lawyer in early settlement negotiations. Typically, these negotiations take place at a time when many facts are still in dispute and liability has not been admitted or clearly established. Also, the employer or its insurance carrier typically is paying for the misconduct alleged to have been committed by an individual, and the "punishment" is inflicted upon the payor, not the harasser himself. Accordingly, the employer may be trying to avoid the negative publicity and damage to its reputation that may come from the lawsuit or claim, even though the wrongdoer may already have been disciplined or fired. Subjecting the employer to public disclosure of the claim or the settlement may serve as a disincentive for the employer to settle the lawsuit for what the plaintiff would consider to be a reasonable amount, in part for fear of setting a "floor" for future claims.

In many instances, an employee (or former employee) has a desire to keep her claims confidential for fear of reprisal or recrimination in future personal or business relationships. For some private individuals (as opposed to celebrities or other public figures), making their claims known to the perpetrator and his or her employer is enough to satisfy the individuals, and they have no need to publicly shame the wrongdoer or the company or publicize their claims.

For most employees, payment of damages, an apology or explanation, and a commitment by the employer to make changes to ensure that the misconduct will not be repeated are sufficient compensation, and they have no need to publicize their settlement — especially if this means they will get higher damages in exchange for keeping the settlement confidential.

But the #MeToo movement has brought about a big change in how companies — and individuals — respond to allegations of sexual misconduct, and today most people (except perhaps public figures and their employers) agree that public figures, especially officials, should not be able to buy their way out of sexual harassment and assault charges through non-disclosure agreements. Yet in the context of civil litigation, there is equal justification for applying the principles of confidentiality that adhere to every other type of case that is resolved through mediation.

It is often the case that neither the employer nor the employee wants the publicity and potential embarrassment that comes from a public airing of this type of very personal experience. In addition, the employer may actually be trying to make things right by paying serious settlement money to the plaintiff

> " … the #MeToo movement has brought about a big change in how companies — and individuals — respond to allegations of sexual misconduct, and today most people (except perhaps public figures and their employers) agree that public figures, especially officials, should not be able to buy their way out of sexual harassment and assault charges through non-disclosure agreements. "

but will not do so willingly if the employer also must suffer the negative publicity from a public disclosure that its employee, especially someone in a leadership role, failed to abide by the employer's own policies. Also, these settlements can readily reach well into six figures, and the business deduction usually realized under corporate tax codes makes that expense much more palatable to a business. Under the Weinstein tax, if the terms of the agreement are kept confidential, there is no deduction available to either side, even for legal fees and costs of litigation.

These types of cases can be expensive, stressful, contentious, and embarrassing to both employer and employee, and the prospect of an early settlement is attractive to both sides as a way to avoid all that. Once the accuser "goes public," however, the employer has less incentive to pay substantial damages because the employer may then want to defend its reputation and protect its "goodwill" in the public eye, which may require harsh litigation and media attention.

Voluntariness, informed decision-making, and self-determination

Mediation is, at its core, a voluntary process. Where there is a willing buyer and a willing seller, mediators usually do not interfere with the bargain the parties strike, even when we believe it may not be in one side's best interest. Other maxims of mediation are that it is designed to give parties the right to make their own decisions and get all the information they need to make well-informed choices. As long as mediators believe that one party has not been coerced or overpowered, once again we usually do not interfere with the parties' decision.

But informed decision-making has limits: When a mediator uses private caucuses, this essentially guarantees an imperfect or incomplete exchange of information, as the mediator will not reveal all of what either side says or believes to the other. Consequently, there are limitations on the "informed" aspect of informed decision-making when parties divulge their confidential interests or rationale for settlement to the mediator in private caucus.

If sexual harassment or abuse cases are uniquely carved out as an exception to the general confidentiality protections afforded in mediation, the parties may be less willing to disclose the underlying interests

> **If sexual harassment or abuse cases are uniquely carved out as an exception to the general confidentiality protections afforded in mediation, the parties may be less willing to disclose the underlying interests or motivations toward settlement.**

or motivations toward settlement. In other words, they may not be making a decision out of their own free will, but rather the optics of how the settlement will appear in the public eye.

Some options to consider

When working on sexual harassment cases, plenty of creative options may serve to maximize recovery, minimize tax consequences, and preserve the integrity of the confidential mediation process. Where both sides are prepared to carefully consider the consequences of a confidentiality agreement or an express disclaimer of confidentiality, settlements may be more achievable and enforceable than they would be if people simply gave in to the new wave of transparency without thinking about everyone's best interests.

In cases where both sides want the settlement to be confidential, mediators and parties still have several good options to consider in the course of the negotiations. The parties can explicitly allocate a reasonable amount of the damages to claims arising out of sexual harassment, leaving the balance of the payment of damages for other, non-sexual harassment-related claims, such as wage and hour violations, retaliation for whistleblowing, disability discrimination, or whatever else has been raised in the pleadings or other demand letters. Those damages will be non-confidential and therefore can still be deducted for tax purposes. The amount must be fair and reasonable, to avoid running afoul of the new tax law, but the other contentions or allegations don't need to be the predominant cause where the claims

are broader than a single cause of action of sexual harassment or abuse.

Or the parties can negotiate the terms of the confidentiality and reflect that in their agreement so that it does not rise to the level of a "non-disclosure agreement." For example, the claims and terms of agreement may be disclosed "upon request" by subpoena or in the course of other legal processes but may not be subject to general disclosure via media or other private communication except to a spouse, attorney, or accountant.

As a third option, the parties can cooperate in drafting an approved statement that will constitute the public disclosure if either party is asked. For example, specific language could state "The parties to this lawsuit have decided it is in both side's best interests to resolve the pending dispute in order to focus upon business and personal matters. Accordingly, effective immediately, employee has dismissed her claims against the employer and any further inquiries should be directed to the human resources director."

Finally, the parties can expressly expunge all preliminary non-disclosure agreements but maintain that the terms of the settlement will not be publicized without notice to the company in advance — and if the terms are made public, provide for an opportunity to craft an acceptable statement to release to current employees and to the public.

Conclusion

There are no easy answers to questions such as whether it's justifiable to allow people in cases of race discrimination — but not those involved in sexual harassment and misconduct cases, for example — to

keep details of their claims confidential. However, there are practical ways to address the parties' desire to maintain confidentiality while still taking advantage of the tax benefits that are afforded to all other employment-related cases, by allocating only a part of the settlement to the sexual harassment-based claims. As with all matters in mediation, the agreement to disclose or keep information confidential is a voluntary matter that is subject to negotiation between the parties, helping them arrive at the best possible solution. Because mediation is meant to be self-determinative, the skillful mediator will know to raise the issue but not provide the solution.

In the end, only after the parties carefully analyze the costs and benefits of confidentiality and the public's right to know can they determine whether they want to keep the terms of the agreement confidential. New tax laws make confidentiality less financially appealing, but there are still plenty of justifications to keep mediated agreements in such cases confidential, just as there are in other cases.

And what about a mediator's moral duty to protect future victims and not contribute to "hushing" up a culture of harassment? The reality is that in today's world, allegations against public figures, celebrities, and high-profile individuals are often likely to come to light, helping inform the public, comfort other victims, and ensure that sexual misconduct is well understood and never tolerated. Mediators can sleep soundly knowing that parties who choose mediation, rather than trying their case in the courtroom or the public forum, typically do so with good reason. And with understanding of various confidentiality options, mediators can protect and safeguard the dispute's details as well as its eventual resolution. ■

Endnotes

1 Cal. Evid. Code Section 1119.

Jan Frankel Schau *is a neutral with ADR Services in Southern California and an adjunct faculty member at the Straus Institute of Dispute Resolution. She focuses on mediating cases arising out of employment, business, real estate, and tort disputes. A Distinguished Fellow of the International Academy of Mediators and former President of the Southern California Mediation Association, she practiced as an employment litigator for more than 20 years. She can be reached at JFSchau@adrservices.com.*

VENUS, MARS, AND THE LAW: ON MEDIATION OF SEXUAL HARASSMENT CASES*

BARBARA J. GAZELEY**

I. INTRODUCTION

Over thirty years ago, Congress enacted the Civil Rights Act of 1964, with its inclusion of "sex" as a prohibited basis of discrimination. Since that time, social mores and legal standards of acceptable workplace behavior have changed dramatically. Yet, sexual harassment in the workplace remains a serious problem. In the words of one lawyer: "I have been handling sexual harassment cases for more than a decade. Just when I think I can no longer be shocked, I am shocked. . . . Just when I think things are better, I find irrefutable evidence that they are not."[1] Statistics on the prevalence of sexual harassment vary, but the most reliable studies indicate continuing harassment of large numbers of women. The most often-cited data come from a well-designed, random survey of 24,000 United States government employees, which was conducted by the Merit Systems Protection Board in 1981 and updated in 1988 and 1995.[2] In both 1981 and 1988, responses indicated that 42% of the women and 15% of the men surveyed had experienced sexual harassment in the preceding two years.[3] Although some women

---

1. Roxanne Barton Conlin, *Opening Our Eyes to Sexual Harassment*, 29 TRIAL, May 1993, at 7, 7.

2. U.S. MERIT SYSTEMS PROTECTION BOARD, SEXUAL HARASSMENT IN THE FEDERAL WORKPLACE: IS IT A PROBLEM? (1981) [hereinafter 1981]; U.S. MERIT SYSTEMS PROTECTION BOARD, SEXUAL HARASSMENT IN THE FEDERAL WORKPLACE: AN UPDATE (1988) [hereinafter 1988]; U.S. MERIT SYSTEMS PROTECTION BOARD, SEXUAL HARASSMENT IN THE FEDERAL WORKPLACE: TRENDS, PROGRESS, CONTINUING CHALLENGES (1995) [hereinafter 1995].

3. *See supra* note 2. *See also* Ellison v. Brady, 924 F.2d 872, 880 n.15 (9th Cir.

harass, it has been estimated that 90-95% of harassers are male.[4] The 1995 report of the Merit Systems Protection Board reveals that many people still do not recognize some types of illegal behavior as sexual harassment.[5] The same report indicates that many people believe that a harasser's intent determines whether the conduct is unlawful sexual harassment.[6]

Although 95% of sexual harassment cases settle before trial,[7] a significant percentage of cases that go to trial result in a plaintiff's verdict.[8] Damages can be substantial: Some verdicts have exceeded the million-dollar mark,[9] and the average jury verdict from 1988 through 1992 was reported at $181,847.[10] The

---

1991). Other surveys indicate that 50-90% of American women are likely to be victims of sexual harassment at some time during their work lives, and that in 55% of cases where the harassee is male, the harasser also is male. *See, e.g.*, Dara A. Charney et al., *An Overview of Sexual Harassment*, 151 AM. J. PSYCHIATRY 10 (1994); Linda Stamato, *Sexual Harassment in the Workplace: Is Mediation an Appropriate Forum?*, 10 MEDIATION Q. 167 (1992) [hereinafter Stamato, *Sexual Harassment*]; Linda Stamato, *The Case for Mediating Sexual Harassment*, SUPP. N.J. L.J., Aug. 15, 1994, at 16, 16 [hereinafter Stamato, *The Case for Mediating*].

4. Andrea Williams, *Model Procedures for Sexual Harassment Claims*, ARB. J., Sept. 1993, at 66, 74 n.7 (citing Louise Fitzgerald et al., *The Incidence and Dimensions of Sexual Harassment in Academia and the Workplace*, 32 J. VOC. BEHAV. 152, 152-72 (1988)). *See also* M. Komaromy et al., *Sexual Harassment in Medical Training*, 328 NEW ENG. J. MED. 322, 326 (cited in Charney et al., *supra* note 3, at 17 n.30).

5. *See* 1995, *supra* note 2, at 6.

6. *Id.* One commentator has noted that these results "support the notion that people disciplined for harassment may tend to feel wronged and may want to be vindicated." Alba Conte, *When the Tables Are Turned*, 32 TRIAL, March 1996, at 30, 31. Another writer has commented that "unlike Justice Potter Stewart's view of pornography, sexual harassment is not immediately recognized by those who see it, or even by those who experience it." Williams, *supra* note 4, at 66.

7. Robert J. Aalberts & Lorne Seidman, *Seeking a "Safe Harbor": The Viability of Summary Judgment in Post-Harris Sexual Harassment Litigation*, 20 S. ILL. U. L.J. 223, 240 (1996) (citing JAMES N. DERTOUZOS & LYNN A. KAROLY, LABOR-MARKET RESPONSES TO EMPLOYER LIABILITY 36 (1992)).

8. Statistics cited for plaintiff verdicts range from 33-64%. *See* 1981, *supra* note 2, at 81 (50%); Stamato, *Sexual Harassment*, *supra* note 3, at 167 (64%); D.E. Terpstra & D.D. Baker, *Outcomes of Sexual Harassment Charges*, 31 ACAD. MGMT. J. 185, 191 (1988) (one-third).

9. *See* Mary Maloney Roberts, *Costly Lessons: What Have Firms Learned from the "Baker" Verdict?*, L.A. DAILY J., Feb. 9, 1995, at 6 (taking note of the $7.1 million jury verdict in *Rena Weeks v. Baker & McKenzie and Martin Greenstein*). The award included $6.9 in punitive damages, later reduced by the court to $3.5 million, which was what the plaintiff had sought. *Id.*

10. Christine Woolsey, *Employers Review Harassment Policies*, BUS. INS., Nov. 15, 1993, at 1, 1 (cited in Aalberts & Seidman, *supra* note 7, at 224). Another article reported the average jury verdict at $460,000, without citing any supporting studies. Stamato, *The Case for Mediating*, *supra* note 3, at 16. Conversely, another article has

average defense costs were reported at $80,000.[11] However, the complete cost of litigation in sexual harassment cases can be as high as $200,000, depending on the case's length and complexity.[12] The risks posed by sexual harassment cases are greater now than ever before[13] because of changes in social and legal standards, as well as the unpredictability of jury behavior in cases where credibility decisions control the case's outcome. Due to the risks, cost, emotional trauma, and time consumed by litigation, lawyers and parties increasingly are choosing to attempt to resolve these disputes in mediation.[14]

There is a minority view that mediation is never appropriate in a sexual harassment case because of the power imbalance between the "victim" and "perpetrator." Mori Irvine, in *Mediation: Is It appropriate for Sexual Harassment Grievances?*, articulates this argument well.[15] It should be noted that this view is driven by a perception that sexual harassment cases are strongly analogous to cases involving criminal assault, rape, or domestic violence because they uniformly involve a severe imbalance of power, rendering the woman[16] incapable of participating effectively in mediation.[17] In the Irvine article, there is an emphasis on "punishment," rather than on resolving the dispute in a positive way.[18] Irvine feels that punishment is necessary in order to "send the message" to other potential harassers that so-

---

reported that "when plaintiffs do prevail, settlements have typically been quite small." Louise F. Fitzgerald et al., *Why Didn't She Just Report Him? The Psychological and Legal Implications of Women's Responses to Sexual Harassment*, 51 J. Soc. Issues 117, 123 (1995) (citing Terpstra & Baker, *supra* note 8).

11. *See* Stamato, *Sexual Harassment, supra* note 3, at 167.

12. Aalberts & Seidman, *supra* note 7, at 229.

13. Roberts, *supra* note 9, at 6.

14. Some commentators have suggested that lawyers "do their clients a serious disservice" and perhaps "should be liable for malpractice" if they counsel their clients to litigate without first trying mediation. Edward J. Costello, Jr., *The Mediation Alternative in Sexual Harassment Cases*, ARB. J., March 1992, at 16, 20 n.12 (quoting Professor Robert F. Cochran, Jr.); Sara Adler, *Sexual Harassment Claims Lend Themselves to Mediation*, L.A. DAILY J., Feb. 18, 1994, at 7.

15. Mori Irvine, *Mediation: Is It Appropriate for Sexual Harassment Grievances?*, 9 OHIO ST. J. ON DISP. RESOL. 27 (1993).

16. As previously mentioned, men may be the victims of sexual harassment. However, because most victims are women, and for purposes of readability, this Article uses the term "woman" in place of "victim." *See supra* note 3 and accompanying text.

17. Irvine, *supra* note 15, at 28.

18. *Id.* at 50.

ciety will not tolerate such behavior.[19] While that may be true in
some cases, it also is true that in many cases an educative ap-
proach, which restores both parties' dignity, can be much more
satisfying to all concerned.

Mediation has been demonstrated to be an effective way of
resolving many sexual harassment cases,[20] and studies show that
most women are more concerned with making the problematic
conduct stop and getting on with their jobs and careers than they
are with blaming and punishing the harasser.[21] Therefore, it ap-
pears that while mediators and attorneys should be alert for
power balance issues, and should either take action to address
problematic power imbalances or refuse to mediate in such
cases, many, and perhaps most, women are sufficiently "empow-
ered" to profit from mediation. Indeed, mediation has been
shown to be an empowering process because it provides an op-
portunity for the woman who complains of harassment to "tell
her story" in a tolerant and respectful environment. Experience
has shown that mediation of sexual harassment cases often pro-
duces greater satisfaction for all parties than do other more ad-
versarial approaches.[22]

In order to provide context, the next Part of this Article re-
views the history of changes in the law and social standards re-
lated to sexual harassment. The third Part discusses general
differences in perception between men and women regarding
sexual harassment behaviors. These differences tend to help cre-
ate disparity in how men and women recall specific events, and
affect their interpretations of behavior. The fourth Part de-
scribes mediation and addresses the advantages of resolving
many sexual harassment disputes through mediation rather than
arbitration or litigation. The last Part discusses factors and strat-

---

19. *Id.*

20. *See, e.g.,* Costello, *supra* note 14; Stamato, *The Case for Mediating, supra* note
3; Williams, *supra* note 4; Anthony Aarons, *Mediation Good Forum for "Delicate" Top-
ics,* L.A. DAILY J., May 17, 1993, at S33; Adler, *supra* note 14; Norman Brand, *Resolv-
ing Sexual Harassment Complaints in the Workplace,* L.A. DAILY J., June 4, 1993, at 7.

21. Charney et al., *supra* note 3, at 13. *See also* Williams, *supra* note 4, at 68.

22. The "experience" referred to is that of this author, of other mediators, of those
persons listed in the acknowledgment at the beginning of this Article (note **) and of
the authors of articles and texts cited herein. *See also, e.g.,* Amy Saltzman, *Life After
the Lawsuit,* U.S. NEWS & WORLD REP., Aug. 19, 1996, at 57 (relating the difficulties
involved in sexual harassment litigation and ongoing struggles in the lives of four wo-
men who sued their employers and won); Andrea Bernstein, *Sex Harassment Suits: The
Fight for Damages Gets Uglier,* Ms., July-Aug. 1996, at 18.

egies that may affect the success of a sexual harassment mediation.

II. A Socio-Legal History of Sexual Harassment

In little more than thirty years, our society has seen a complete upheaval in social customs and legal standards regarding both the employment of women and the social and sexual behavior of men and women. In the 1950s and 1960s, relatively few women worked in male-dominated professions and trades. For the most part, jobs for women were limited to a few acceptable types of employment, such as secretarial work, waitressing, teaching, and nursing. In the early 1970s, women moved into professions and trades in increasingly larger numbers. This caused great change in male and female gender identity, standards of acceptable behavior, and legal consequences attached to behavior. The social and legal history that follows demonstrates how greatly times have changed. Retaining an awareness of this legal history may facilitate the development of empathy for the real people involved in sexual harassment cases.

To some extent, a clash of cultures continues to exist in the workplace: Older versus newer mores and legal requirements, male versus female perspectives. For example, an older man may be "clueless" about how his behavior affects the women around him, while a younger woman may not be aware that her openness about sex and her casual dress may be misinterpreted by men raised in earlier times. Similarly, an older woman may be more sensitive to sexual innuendo and jokes than a younger woman might be, and men may not understand that a woman who enjoys sexual humor might not enjoy sex jokes that demean women, and conversely, that women who object to demeaning humor are not necessarily oversensitive "prudes." Further, men may respond inappropriately to the insecurity they may feel as a result of the influx of women into formerly male-dominated areas.

Occasionally, claimants and defendants lie about what actually happened in the workplace, and sometimes harassment is intentional or malicious. Much more frequently, the "truth" is less clear. Research shows that men and women have different perceptions of what type of behavior constitutes sexual harassment, and different interpretations of men's motivations for

problematic behavior.[23] A man accused of harassment may feel genuinely affronted because he intended no harm. Further, he may deny the charge of "harassment" because he didn't intend to harass; research shows that many people still believe that harassment requires bad intent.[24] In the same situation, the woman may be justifiably aggrieved by the man's inappropriate conduct, regardless of his motivation; she may see an abuse of power of which the man is unaware. Conceding the view that harassing behavior never has been appropriate, it is nonetheless true that cultural standards have changed over the past thirty years, and that these changes have contributed to confusion in both men and women about how they should interact in the workplace. Maintaining a sense of how dramatically customs and legal standards have changed, and still are changing, is useful to those who attempt to resolve sexual harassment disputes in a less traumatic way.

A. *Title VII and Early Cases: The "Denial" Phase*

Under Title VII of the Civil Rights Act of 1964, it is an unlawful employment practice for an employer to "discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin."[25] The term "sex" was added at the last minute on the floor of the House of Representatives, and there is little legislative history to provide evidence of congressional intent.[26] There has been some speculation that those attempting to defeat Title VII proposed the amendment, on the theory that the prohibition of employment discrimination against women was so laughable that it would torpedo the entire effort.[27]

---

23. *See infra* Part III. *See also* Carol A. Ford & Francisco J. Donis, *The Relationship Between Age and Gender in Workers' Attitudes Toward Sexual Harassment*, 130 J. PSYCHOL. 627 (1996).

24. *See* 1995, *supra* note 2, at 6.

25. 42 U.S.C. § 2000e-2 (1994).

26. 110 CONG. REC. 2577-84 (1964).

27. Jo Annette Jacobs, *No More Nervous Breakdowns: Sexual Harassment and the Hostile Work Environment*, 62 U. MO. KAN. CITY L. REV. 521, 528 n.57 (1994) ("The amendment was introduced by Representative Smith of Virginia and most of the declarations in support of [the amendment] were either offered by female representatives or men who represented southern states. . . . Indeed, as was aptly pointed out by the Oregon representative Mrs. Green, many who voiced support for the addition of 'sex' had

In the mid-1970s, sexual harassment cases began to percolate through the courts. Some early cases reflected the social mores of the time, as well as the predisposition of some older, conservative male judges, who viewed "sexual harassment" as merely a fact of life that women must accept if they wished to remain in their chosen occupation. For example, in *Corne v. Bausch & Lomb*, two women claimed that they had been repeatedly subjected to their supervisor's verbal and physical sexual advances.[28] The case was dismissed for failure to state a claim under Title VII.[29] The court found that the supervisor's alleged behavior was merely a "personal proclivity, peculiarity, or mannerism" and that he had been "satisfying a personal urge."[30] The conduct was not a company policy, and the company did not benefit from the conduct; therefore, it could not be considered a "company directed policy which deprived women of employment opportunities."[31] The court also expressed concern that "an outgrowth of holding such activity to be actionable under Title VII would be a potential federal lawsuit every time any employee made amorous or sexually oriented advances toward another. The only sure way an employer could avoid such charges would be to have employees who were asexual."[32]

B.   The Equal Employment Opportunity Commission (EEOC) Guidelines on Discrimination Because of Sex

In 1980, the EEOC issued its Guidelines on Discrimination Because of Sex, which state:

---

opposed the passage of the Equal Pay Act of 1963, just a year earlier."). *See also* Deanna Weisse Turner, Note, *Civil Rights—Employer's [sic] Beware: The Supreme Court's Rejection of the Psychological Injury Requirement in* Harris v. Forklift Systems, Inc., 114 S. Ct. 367 (1993), *Makes It Easier for Employees to Establish a Claim for Sexual Harassment Based on a Hostile Working Environment*, 17 U. ARK. LITTLE ROCK L.J. 839, 843-44, n.47 (1995) (noting that southern men were concerned about the possibility of "colored" or "Negro" women getting better treatment than "white Christian" women).

28. 390 F. Supp. 161 (D. Ariz. 1975), *vacated by* Corne v. Bausch & Lomb, Inc., 562 F.2d 55 (9th Cir. 1977).

29. *Id.* at 162.

30. *Id.* at 163.

31. *Id.*

32. *Id.* at 163-64. *See also* Barnes v. Train, 13 FAIR EMPL. PRAC. CAS. (BNA) 123, 124 (D.D.C. 1974), *rev'd sub nom.* Barnes v. Costle, 561 F.2d 983 (D.C. Cir. 1977) (at the administrative level, discrimination was found to be based not on the plaintiff's sex, but on her refusal to have sex with her supervisor, which resulted in an "inharmonious personal relationship.").

> Harassment on the basis of sex is a violation of section 703 of
> title VII. Unwelcome sexual advances, requests for sexual
> favors, and other verbal or physical conduct of a sexual na-
> ture constitute sexual harassment when (1) submission to
> such conduct is made either explicitly or implicitly a term or
> condition of an individual's employment, (2) submission to or
> rejection of such conduct by an individual is used as the basis
> for employment decisions affecting such individual, or
> (3) such conduct has the purpose or effect of unreasonably
> interfering with an individual's work performance or creating
> an intimidating, hostile, or offensive working environment.[33]

The first two subsections above set out what has become the test
for "quid pro quo" sexual harassment. The third subsection de-
scribes "hostile environment" harassment. The Guidelines fur-
ther provide that a determination of sexual harassment will be
made by looking at "the record as a whole and at the totality of
the circumstances, such as the nature of the sexual advances and
the context in which [they] occurred."[34] An employer is to be
responsible for its acts

> and those of its agents and supervisory employees . . . regard-
> less of whether the specific acts complained of were author-
> ized or even forbidden by the employer and regardless of
> whether the employer knew or should have known of their
> occurrence . . . . With respect to conduct between fellow em-
> ployees, an employer is responsible for acts of sexual harass-
> ment in the workplace where the employer (or its agents or
> supervisory employees) knows or should have known of the
> conduct, unless it can show that it took immediate and appro-
> priate corrective action.[35]

The Guidelines also state that the employer can be held respon-
sible for the acts of nonemployees (such as clients or customers)
under the same "knew or should have known" standard.[36]

C. *"Quid Pro Quo" and the "Hostile Environment"*

The first cases in which the courts granted relief, in the late
1970s and early 1980s, distinguished between mere "individual"
or "personal" sexual advances (which the courts were reluctant
to find illegal) and those in which the granting of sexual favors

---

33. 29 C.F.R. § 1604.11(a) (1980).
34. *Id.* § 1604.11(b).
35. *Id.* § 1604.11(c), (d).
36. *Id.* § 1604.11(e).

was made a condition of continued employment, promotion, or other job benefits. In *Henson v. City of Dundee*,[37] the court stated: "An employer may not require sexual consideration from an employee as a *quid pro quo* for job benefits."[38] The employer is strictly liable for quid pro quo sexual harassment.[39]

"Hostile environment" claims were sustained first in *Rogers v. EEOC*,[40] a 1971 case of discrimination based on national origin. The court held that "the phrase 'terms, conditions and privileges of employment' in [Title VII] is an expansive concept which sweeps within its protective ambit the practice of creating a working environment heavily charged with ethnic or racial discrimination."[41] A few years later, feminist legal scholar Catherine MacKinnon authored a text asserting that sexual harassment also creates a hostile or abusive work environment and impermissibly alters the conditions of women's employment, thereby violating Title VII.[42] However, it was not until 1986 that the United States Supreme Court addressed this issue and concluded that "hostile environment" sexual harassment provides a basis for relief under Title VII.[43] In *Meritor*, the Court held that "a plaintiff may establish a violation of Title VII by proving that discrimination based on sex has created a hostile or abusive work environment."[44]

In *Meritor*, plaintiff Vinson claimed that her supervisor, Taylor, made demands for sex, which she initially refused but eventually accepted because she was afraid of losing her job.[45] Vinson claimed that Taylor's actions spanned three years, and included forty to fifty acts of forced sexual intercourse, fondling her in front of other employees, exposing himself, following her

---

37. 682 F.2d 897, 908 (11th Cir. 1982), *superceded by statute*, WASH. REV. CODE ANN. § 49.60.180 (West 1990 & Supp. 1993).

38. *Id. See also* Miller v. Bank of America, 600 F.2d 211 (9th Cir. 1979); Barnes v. Costle, 561 F.2d 983 (D.C. Cir. 1977); Tomkins v. Public Serv. Elec. & Gas Co., 568 F.2d 1044 (3d Cir. 1977) (finding a strong congressional mandate for providing relief, and dismissing concerns based on judicial economy); Williams v. Saxbe, 413 F. Supp. 654 (D.D.C. 1976).

39. *See, e.g.*, Nichols v. Frank, 42 F.3d 503 (9th Cir. 1994).

40. 454 F.2d 234 (5th Cir. 1971).

41. *Id.* at 238.

42. CATHERINE MACKINNON, SEXUAL HARASSMENT OF WORKING WOMEN (1979).

43. Meritor Sav. Bank v. Vinson, 477 U.S. 57 (1986).

44. *Id.* at 66.

45. *Id.* at 60.

into the rest room, and forcible rape.[46] These activities ceased only after the plaintiff started going with a steady boyfriend.[47] Vinson testified that because she was afraid of Taylor, she never reported his harassment to any of his supervisors and never attempted to use the bank's complaint procedure.[48] Taylor denied Vinson's allegations, saying that she made them up in response to a business dispute.[49] The district court denied relief, but did not resolve the conflicting testimony.[50] Instead, it found that Vinson's participation was "voluntary" and "had nothing to do with her continued employment . . . or her advancement or promotions at [the bank]."[51] The court found that Vinson had not been sexually harassed and further found that, because of Vinson's failure to lodge a complaint, the bank was without notice and could not be held liable for Taylor's actions.[52] The circuit court reversed, in reliance on the EEOC guidelines, and found that Vinson's harassment was of the "hostile environment" variety.[53] The circuit court found that Vinson's "voluntariness" was immaterial if toleration of sexual harassment was a condition of her employment, and held that an employer is absolutely liable for harassment practiced by supervisory personnel, regardless of whether the employer knew or should have known of the misconduct.[54]

The Supreme Court granted certiorari, ultimately affirming in part on other grounds.[55] The employer argued that Title VII's protection was limited to tangible or economic loss and did not apply to "purely psychological aspects of the workplace environment."[56] The Supreme Court rejected that defense, relying on the EEOC's Guidelines Regarding Sexual Harassment, which permit sexual harassment claims based on an "intimidating, hostile or offensive working environment."[57] Where the trial court

---

46. *Id.*
47. *Id.*
48. *Id.* at 61.
49. *Id.*
50. *Id.*
51. *Id.*
52. *Id.*
53. *Id.*
54. *Id.* at 62-63.
55. *Id.* at 63.
56. *Id.* at 64.
57. *Id.* at 65.

had focused on Vinson's "voluntary" participation and found that the relationship had not affected her employment, the Supreme Court emphasized the "unwelcome" nature of Taylor's conduct.[58] The Court held that a plaintiff may recover for "hostile environment" sexual harassment by demonstrating that unwelcome sexual conduct was severe or pervasive enough to alter the conditions of employment and create an abusive work environment.[59] The Court refused to apply strict liability in hostile environment cases, but held that employers could be liable for supervisors' acts under agency principles, which it declined to clarify.[60] The Court also found that the mere existence of a sexual harassment policy could not insulate the employer from liability, and refused to require proof of "tangible" or "economic" injury in hostile environment cases.[61]

After *Meritor*, a split in the circuits developed over the degree of harm required to prove a hostile environment sexual harassment case. In *Meritor*, the Court referred to an "abusive" environment and to the victim's "emotional and psychological stability," which led some courts to impose a requirement of psychological harm.[62]

In *Rabidue v. Osceola Refining Co.*,[63] the plaintiff alleged that she and other female employees were subjected to obscene comments by a supervisor, that male employees displayed pictures of naked women, and that she was undermined in a number of other ways. In that case, the court held plaintiff had to prove that the hostile environment seriously affected her psy-

---

58. *Id.* at 68.
59. *Id.* at 67.
60. *Id.* at 72. The Court stated:
We therefore decline the parties' invitation to issue a definitive rule on employer liability, but we do agree with the EEOC that Congress wanted courts to look to agency principles for guidance in this area. While such common-law principles may not be transferable in all their particulars to Title VII, Congress' decision to define "employer" to include any "agent" of an employer . . . surely evinces an intent to place some limits on the acts of employees for which employers under Title VII are to be held responsible. For this reason, we hold that the Court of Appeals erred in concluding that employers are always automatically liable for sexual harassment by their supervisors. For the same reason, absence of notice to an employer does not necessarily insulate that employer from liability.
Id. (citations omitted).
61. *Id.* at 64.
62. *Id.* at 60.
63. 805 F.2d 611 (6th Cir. 1986).

chological well-being.[64] The court found that the obscene remarks and pictures, while annoying, did not cause psychological harm, and declined to impose liability.[65]

Similarly, in *Sparks v. Pilot Freight Carriers, Inc.*,[66] *Downes v. FAA*,[67] and *Paroline v. Unisys Corp.*,[68] the courts found that in order for the harassment to be sufficiently severe to support liability, it must affect the plaintiff's psychological well-being seriously. Conversely, in *Ellison v. Brady*,[69] a case in which the harasser exhibited obsessive, stalking-type behavior toward a coworker whom he barely knew, the court specifically disagreed with the "psychological harm" requirement of *Rabidue* and declined to impose it. The Ninth Circuit found that *Meritor* did not dictate such a standard and stated that "it is the harasser's conduct which must be pervasive or severe, not the alteration in the conditions of employment. Surely, employees need not endure sexual harassment until their psychological well-being is seriously affected to the extent that they suffer anxiety and debilitation."[70] The Ninth Circuit in *Ellison* adopted a "reasonable

---

64. *Id.*

65. The court also quoted with approval the language of the district court opinion, which stated:

> Indeed, it cannot seriously be disputed that in some work environments, humor and language are rough hewn and vulgar. Sexual jokes, sexual conversations and girlie magazines may abound. Title VII was not meant to—or can—change this .... Title VII [was not] designed to bring about a magical transformation in the social mores of American workers.

Id. at 620-21. A strong dissenting opinion noted that "the majority suggests . . . that a woman assumes the risk of working in an abusive, anti-female environment . . . [and] contends that such work environments somehow have an innate right to perpetuation . . . ." *Id.* at 626 (Keith, J., concurring in part, dissenting in part). Judge Keith described many more egregious details of the harassing conduct, including the fact that the plaintiff's supervisor referred to women as "whores," "cunt," "pussy," and "tits"; that a poster on the wall in the workplace "showed a prone woman who had a golf ball on her breasts with a man standing over her, golf club in hand, yelling 'Fore'"; that the supervisor prevented plaintiff from visiting or taking customers to lunch as all previous male credit managers had done; that the plaintiff was denied the "perks" that were accorded to male management employees; and that the supervisor in other ways undermined plaintiff and made it difficult for her to do her job. He found that the supervisor's conduct evinced a clear "anti-female animus," which should be imputed to the employer under agency principles, per *Meritor*. *Id.* at 624-25.

66. 830 F.2d 1554 (11th Cir. 1987).

67. 775 F.2d 288 (Fed. Cir. 1985).

68. 879 F.2d 100 (4th Cir. 1989), *vacated in part*, 900 F.2d 27 (4th Cir. 1990).

69. 924 F.2d 872 (9th Cir. 1991).

70. *Id.* at 878.

woman" standard,[71] which the court said would classify conduct as unlawful sexual harassment "even when harassers do not realize that their conduct creates a hostile working environment."[72] The court also held that the employer, which had initially transferred the harasser to a different office but returned him when he filed a grievance action, did not take strong enough remedial action.[73]

D. Our Awakening: Anita Hill and Clarence Thomas

In October 1991, men and women watched the Anita Hill/ Clarence Thomas drama unfold on national television. During the Senate hearings regarding Thomas' nomination to the Supreme Court, Hill accused Thomas of sexually harassing her ten years earlier when he was her boss at the EEOC. Thomas vehemently denied her claims, and stated that the alleged harassment incidents (comments about pubic hairs on Coke cans, sexual innuendo, lewd comments, and unwanted advances and requests for dates) simply did not happen. Both spoke with great conviction. Much was made of the fact that Hill did not complain of the incidents at the time they happened; little thought was given to the possibility that workplace standards had changed in the preceding ten years, and that Thomas' behav-

---

71. *Id.* at 878-79. The court stated that:

Conduct that many men consider unobjectionable may offend many women. . . . *See also* Ehrenreich, *Pluralist Myths and Powerless Men: The Ideology of Reasonableness in Sexual Harassment Law* . . . (men tend to view some forms of sexual harassment as "harmless social interactions to which only overly-sensitive women would object") . . . [W]e hold that a female plaintiff states a prima facie case of hostile environment sexual harassment when she alleges conduct which a reasonable woman would consider sufficiently severe or pervasive to alter the conditions of employment and create an abusive working environment . . . [W]e believe that a sex-blind reasonable person standard tends to be male-biased and tends to systematically ignore the experiences of women.

Id. (citations omitted).

72. *Id.* at 880.

73. *Id.* at 882.

The record on appeal suggests that Ellison's employer did not express strong disapproval of Gray's conduct, did not reprimand Gray, did not put him on probation, and did not inform him that repeated harassment would result in suspension or termination. Apparently, Gray's employer only told him to stop harassing Ellison. Title VII requires more than a mere request to refrain from discriminatory conduct. Employers send the wrong message to potential harassers when they do not discipline employees for sexual harassment.

Id. (citations omitted).

ior in the early 1980s was being judged by early 1990s standards.[74]

The debate during the hearings, and in the weeks and months that followed, centered on a search for the objective truth, much like a civil trial. The central question was: Who lied? Was Thomas denying the truth, or was Hill fabricating stories to get back at a boss she disliked?[75] Few seemed to consider the possibility that both were telling the truth, from their differing perspectives: that incidents so agonizingly embarrassing and painful to Hill that they created permanent memories were so trivial and inconsequential to Thomas that he had forgotten them completely.[76]

Polls taken at the close of the hearings showed that a majority of both men and women believed Thomas, although more women than men accepted Hill's version of the events.[77] However, follow-up polls conducted a year after the hearings produced the opposite result: A majority of both men and women now believed Hill, although more women than men did so.[78] Commentators theorized that in the intervening year, men and women talked to each other and discussed women's harassment experiences, in many cases for the first time.[79] One study

---

74. This is not meant as a comment on the "fairness" to Thomas of being judged by different standards than those existing at the time of the alleged harassment; nor is it meant to defend inappropriate behavior. It is merely intended to note the possibility that standards, in fact, had changed and that Thomas may have been caught in the fallout from this cultural evolution.

75. "You clearly have to say one of them is lying," said Democratic Sen. Herb Kohl of Wisconsin. *See* Rhodes Cook, *Hill vs. Thomas*, CONG. Q., Dec. 9, 1995, at 3717, 3719.

76. As a litigator, this author realizes that no two witnesses to an event ever seem to remember the same set of facts. In many cases, from the accounts of eyewitnesses, one might suspect that completely different events were witnessed. At the time of the hearings, the author expressed to friends the possibility that both Hill and Thomas were telling the truth, each from his or her own perspective. Few seemed to find that plausible, however; people's predispositions and the need to find out "the truth" completely overwhelmed the possibility that two legitimately different interpretations might exist. *Cf.* Louise H. Kidder et al., *Recalling Harassment, Reconstructing Experience*, 51 J. Soc. Issues 53, 63 (1995) ("Might the same words be forgotten throw-away lines for him and leave indelible marks on her?"). It also is possible that, over the years, the details became magnified in Hill's mind. Memory is not videotape, after all.

77. See Cook, *supra* note 75, at 3719.

78. Suzanne Garment, *Confirming Anita Hill?*, 4 AM. ENTERPRISE 18 (1993); Dianne Rucinski, *The Polls—A Review: Rush to Judgment? Fast Reaction Polls in the Anita Hill-Clarence Thomas Controversy*, 57 PUB. OPINION Q. 575 (1993).

79. As Roxanne Barton Conlin relates:

showed that women were significantly more likely than men (women 64-71%, men 27-33%) to report "vivid image" or "flash-bulb" memories of sexual harassment as a result of watching the hearings.[80] This may parallel Hill's and Thomas's perceptual differences. In the year following the Hill/Thomas hearings, there was a reported 50-70% increase in the filing of sexual harassment complaints.[81] Thus, the 1991 Hill/Thomas hearings catalyzed a marked shift in men's and women's consciousness of sexual harassment.

E. Putting Some Teeth in Title VII

Coincidentally, in 1991, the same year that the Hill/Thomas hearings took place, Congress amended Title VII, in the Civil Rights Act of 1991.[82] Prior to that time, a plaintiff who won a hostile environment action under Title VII was limited to recovering economic losses, which might include injunctive relief in the form of reinstatement, back pay, lost benefits, attorney's fees, court costs, and interest.[83] Since the 1991 amendment, plaintiffs have had the right to recover additional damages, including: compensatory damages (including those for emotional pain, suffering, inconvenience, mental anguish and loss of enjoyment of life) and punitive damages (where plaintiff must prove malicious intent or reckless indifference to her civil rights).[84]

> One evening, I was discussing the hearings with a group of friends, including a couple who had been married nearly 40 years. The husband asked, "Why didn't she complain? Why didn't she just leave?" His wife turned to him and said, "I didn't when it happened to me." The husband was shocked speechless. To that day, he was unaware that the woman he loved had been sexually harassed in the early years of their marriage. I know that this scenario was repeated again and again all over the country.

Conlin, *supra* note 1, at 7.

80. Claire K. Morse et al., *Gender Differences in Flashbulb Memories Elicited by the Clarence Thomas Hearings*, 133 J. SOC. PSYCHOL. 453 (1993). "Flashbulb" memories are described as vivid, detailed memories of personal experiences of sexual harassment, which many women (and few men) reported having at the time they viewed the Hill/Thomas hearings.

81. Mollie L. Jaschik-Herman et al., *Women's Perceptions and Labeling of Sexual Harassment in Academia Before and After the Hill-Thomas Hearings*, 33 SEX ROLES 439, 440 (1995) (50% increase); *cf.* Charney et al., *supra* note 3, at 13 n.42 (70% increase).

82. Civil Rights Act of 1991, Pub. L. No. 102-166, 105 Stat. 1071 (codified at 42 U.S.C. § 2000e-2 (1994)).

83. Civil Rights Act of 1964, 42 U.S.C. § 2000e-5 (1988) (as amended by Act of Nov. 21, 1991).

84. 42 U.S.C. § 2000e-2 (1994).

Such damages are subject to statutory caps, ranging from $50,000 to $300,000, depending on the size of the company, measured by the number of employees.[85] Plaintiffs suing for compensatory or punitive damages may demand a jury trial and may be awarded expert witness fees, in addition to attorney fees.[86]

F. Clarification: Nervous Breakdowns Not Required

In 1993, the United States Supreme Court in *Harris v. Forklift Systems, Inc.*[87] finally resolved the conflict in the circuits over whether a "hostile environment" case required psychological harm. In that case, the president of Forklift Systems allegedly had made demeaning comments to the plaintiff such as: "You're a woman, what do you know?"; "We need a man as the rental manager"; "Dumb ass woman."[88] He also allegedly threw objects onto the ground and asked her to pick them up, requested that she get change out of his front pocket, and suggested that they discuss her raise at a motel.[89] Further, the plaintiff alleged that he made sexual comments about Harris' and other women's clothing, and implied that Harris succeeded in making a deal with a customer by "promising him some [sex] on Saturday night."[90] The trial court found his conduct offensive but held for the defendant on the basis that the conduct did not affect plaintiff's psychological well-being,[91] and the Sixth Circuit affirmed.[92] The Supreme Court reversed in a unanimous decision written by Justice O'Connor, which was released less than one month after oral argument. The Court found that Title VII "comes into play before the harassing conduct leads to a nervous breakdown" and stated that because "Title VII bars conduct that would seriously affect a reasonable person's psychological well-being . . . there is no need for it also to be psychologically injurious."[93] Reaffirming *Meritor*, Justice O'Connor set out the following "objective-

---

85. *Id.*
86. *Id. See also* Francis Achampong, *Potential Ramifications of the Elimination of the Psychological Harm Requirement in Hostile Environment Sexual Harassment Cases*, 38 How. L.J. 163, 175-76 (1994).
87. 510 U.S. 17 (1993).
88. *Id.* at 19.
89. *Id.*
90. *Id.*
91. *Id.* at 20.
92. *Id.*
93. *Id.* at 22.

subjective" standard: To violate Title VII, the conduct complained of must be "severe or pervasive enough to create an objectively hostile or abusive work environment—an environment that a reasonable person would find hostile or abusive"; in addition, the plaintiff must "subjectively perceive the environment to be abusive."[94]

G.   Unresolved Questions

Sexual harassment law remains a work in process. The current standard has been criticized as vague. However, as Justice O'Connor noted: "This is not, and by its nature cannot be, a mathematically precise test."[95] One commentator has observed that courts have been reluctant to grant summary judgment based on the *Harris* standard, except where they have found that the employer took immediate remedial action to deal with the harassment.[96] Conversely, other exasperated (and notably female) scholars say that it's simple: "If you can't do it or say it in front of your spouse or children, don't do it at the office."[97] The Supreme Court has yet to determine whether the applicable standard should be the "reasonable woman" or the "reasonable person under the same or similar circumstances." Those who have been hotly debating this issue may find relief in a recent study demonstrating that the choice of standard (reasonable per-

---

94. *Id.* at 21-22.

95. *Id.* at 22. See also Justice Scalia's concurring remarks: "I know of no test more faithful to the inherently vague statutory language than the one the Court today adopts. For this reason, I join the opinion of the Court." *Id.* at 25 (Scalia, J., concurring in part). Commentators similarly have discussed the difficulty of determining what behavior constitutes harassment. *See* Eugene Volokh, *Let's Clarify What We Mean by "Sexual Harassment,"* L.A. DAILY J., Jan. 20, 1995, at 6; *see also* Kerry A. Colson, Comment, Harris v. Forklift Systems, Inc.: *The Supreme Court Moves One Step Closer to Establishing a Workable Definition for Hostile Work Environment Sexual Harassment Claims,* 30 NEW ENG. L. REV. 441 (1996).

96. Aalberts & Seidman, *supra* note 7, at 244.

97. Paula Poda, *One Day "No" Might Be Enough: What to Do When Sexual Harassers Just Don't Get It,* BARRISTER, Spring 1993, at 27, 27; *see also* Conlin, *supra* note 1, at 7.

> Those who are confused over the meaning of the law need to remember only a few simple rules. If it is something you would not say or do in front of your mother or sister or daughter, it is wrong to say or do it to a co-worker. If you wouldn't say it to a person of the same sex, why would you need to say it to a person of the opposite sex? If you wouldn't want to see it on the front page of the newspaper, you shouldn't do it. If in doubt, don't.

Id.

son versus reasonable woman or victim) had no effect on the jury verdict.[98] In addition, scholars currently are debating over whether sexual harassment law infringes on First Amendment rights.[99] Newer varieties of claims being asserted include suits by alleged harassers,[100] as well as complaints of third-party harassment by nonemployees, such as clients, customers, and consultants.[101] In addition to the federal claims, sexual harassment cases may be based on analogous state statutes,[102] or common-

---

98. Richard L. Wiener et al., *Social Analytic Investigation of Hostile Work Environments: A Test of the Reasonable Woman Standard,* 19 LAW & HUM. BEHAV. 263, 276 (1995). *See also* Barbara A. Gutek & Maureen O'Connor, *The Empirical Basis for the Reasonable Woman Standard,* J. SOC. ISSUES, Spring 1995, at 151, 162. It is sometimes comforting to know that lawyerly arguments over how many angels can dance on the head of a pin may have no real effect on the jury's determination, which may be a terrific argument for retaining the jury system. For more on the continuing debate, see, for example, Kathryn Abrams, *The Reasonable Woman: Sense and Sensibility in Sexual Harassment Law,* 42 DISSENT 48 (1995); Gillian K. Hadfield, *Rational Women: A Test for Sex-Based Harassment,* 83 CAL. L. REV. 1151 (1995).

Recently, courts and scholars have begun to debate whether the law provides relief for same-gender harassment, which may put a new spin on the "reasonable woman vs. reasonable victim" debate. *See, e.g.,* Renee Levay, *Employment Law: Does/Should Title VII Apply to Same-Gender Sexual Harassment?,* 26 U. MEM. L. REV. 1601 (1996); Julianna Ryan & John M. Butler, *Without Supreme Court Precedent, Federal Courts Struggle with the Issue of Whether Title VII Lawsuits May Be Brought for Same-Sex Sexual Harassment,* NAT'L L. J., Dec. 23, 1996, at B8; Charles Howard Wilson, *Goosing, Bagging and Dry-Humping: Foreplay or Horseplay? Same-Sex Sexual Harassment Claims Under Title VII:* McWilliams v. Fairfax County Board of Supervisors, 21 T. MARSHALL L. REV. 211 (1996); Lisa Fair McEvers, Comment, *Civil Rights—Work Environment: Sexual Harassment: "Sexual Harassment by a Supervisor of the Same Sex, Is It Actionable?",* 72 N.D. L. REV. 397 (1996); Susan Perissinotto Woodhouse, Comment, *Same-Gender Sexual Harassment: Is It Sex Discrimination Under Title VII?,* 36 SANTA CLARA L. REV. 1147 (1996).

99. *See* Wayne Lindsey Robbins, Jr., *When Two Liberal Values Collide in an Era of "Political Correctness": First Amendment Protection as a Check on Speech-Based Title VII Hostile Environment Claims,* 47 BAYLOR L. REV. 789 (1995); *cf.* Deborah Epstein, *Can a "Dumb Ass Woman" Achieve Equality in the Workplace? Running the Gauntlet of Hostile Environment Harassing Speech,* 84 GEO. L.J. 399 (1996); Jacobs, *supra* note 27, at 542; George Rutherglen, *Sexual Harassment: Ideology or Law?,* 18 HARV. J.L. & PUB. POL'Y 487 (1995); Eugene Volokh, *Thinking Ahead About Freedom of Speech and Hostile Work Environment Harassment,* 17 BERKELEY J. EMP. & LAB. L. 304 (1996);

100. Conte, *supra* note 6, at 30.

101. Achampong, *supra* note 86, at 179; *New Legal Worry: Third-Party Sexual Harassment,* HR FOCUS, July 1996, at 1; Lynn Hecht Schafran, *Sexual Harassment: An Ounce of Prevention,* 30 TRIAL 14, 20 (1994); Patricia Konopka, Comment, *Combatting Sexual Harassment in the Workplace Without Risking a Wrongful Discharge Lawsuit: An Employer's Dilemma,* 42 KAN. L. REV. 437 (1994).

102. *See, e.g.,* OR. REV. STAT. § 659.030 (1995). Claims under both the federal and state statutes first must be presented to the appropriate government agency

law theories such as breach of contract, tortious interference with contract, negligence, intentional or negligent infliction of emotional distress, fraud, assault, battery, and invasion of privacy.[103]

Arguably, the standards the Supreme Court set forth in *Meritor* and *Harris* are vague enough that the jury is given only fairly general directions to guide its determination of what constitutes "harassment." Moreover, in many sexual harassment cases, the facts are very much in dispute; the outcome often depends on the jury's credibility determination. Because our legal system is set up as a search for "the truth," the jury must decide which party owns the truth and which party is lying. In many cases, the reality is not that simple. Even in cases in which the facts are undisputed or one side appears to be much more credible than the other, a jury still may be unable to address the parties' "real" interests. Mediation offers the opportunity for creative solutions, as well as greater satisfaction with both the process and the result.

III.   HOW AND WHY MEN "JUST DON'T GET IT": WOMEN AND MEN PERCEIVE SOCIAL AND SEXUAL BEHAVIOR VERY DIFFERENTLY

Over the past twenty years, we slowly have become more aware that men and women simply don't see sexual harassment from the same perspective. This difference in perspective creates the context in which harassment occurs, and tends to cause each side of a sexual harassment dispute to suspect the other side of evil intent. Recent research provides information that can help each side understand the legitimacy of the other's perspective, even where conduct is clearly unacceptable. This understanding may increase the likelihood of apology, forgiveness, and at least the beginnings of healing for both parties. Unlike the adversarial process, mediation offers the chance to educate the participants about their differing perspectives; mediation permits this healing, which cannot be accomplished through monetary awards or defense verdicts. The following is a summary of recent research in this area.

(EEOC/BOLI) for the purpose of obtaining a "right to sue" letter before they may be filed in court.

   103. Williams, *supra* note 4, at 68.

A. Male and Female Differences in Perception of Harassment

Researchers have theorized that "sexual harassment results from the complex interplay of ambivalent motives and gender stereotyping of women and jobs .... Traditionally, men's motivational orientation toward women has been deeply ambivalent, reflecting male desires for both dominance and intimacy."[104] As a result, men may perceive their motivations as positive, even when these motivations encourage conduct that is problematic for women. Thus, discrimination and harassment may be either "benevolent" or "hostile";[105] however, both types are illegal. Men who do not intend to hurt women may be embarrassed and offended when accused of sexual harassment. In addition, a continuum of behavior exists, and both men and women have difficulty applying legal definitions to daily conduct.[106]

Men and women do not agree about what constitutes sexual harassment. Women use broader definitions of sexual harassment, and therefore are more likely to define specific behaviors as harassment.[107] Women also find sexual harassment to be more serious than do men.[108] Men are significantly less likely to perceive behavior as harassing, and are likely to perceive a woman's friendly behavior as a sign of sexual interest and availability.[109] Even though men see male-female exchange as more "sexual" than do women, they also believe that such behaviors are more normative and acceptable than do women.[110] As discussed by the *Ellison* court, this difference may be due to the fact that women have much greater reason to feel threatened by men's sexual behavior because of women's particular vulnerability to rape and violent sexual assault.[111] In addition, women may

---

104. Susan T. Fisk & Peter Glick, *Ambivalence and Stereotypes Cause Sexual Harassment: A Theory with Implications for Organizational Change*, J. Soc. Issues, Spring 1995, at 97, 97-98.

105. *Id.* at 98-99.

106. Williams, *supra* note 4, at 67.

107. Wiener et al., *supra* note 98, at 264.

108. Aron Saperstein et al., *Ideology or Experience: A Study of Sexual Harassment*, 32 Sex Roles 835, 839 (1995).

109. Rick Garlick, *Male and Female Responses to Ambiguous Instructor Behaviors*, 30 Sex Roles 135, 139 (1994).

110. *Id.*

111. Ellison v. Brady, 924 F.2d 872, 879 (9th Cir. 1991). The court stated:

For example, because women are disproportionately victims of rape and sexual assault, women have a stronger incentive to be concerned with sexual behavior. Women who are victims of mild forms of sexual harassment may

perceive, with some basis, that some men view them "less as colleagues and more as objects of sexual challenge."[112]

In a recent study, both sexes tended to agree that the presence of the following factors indicated that sexual harassment had taken place: Quid pro quo harassment, unwanted physical contact, and superior/subordinate relationships.[113] Despite this common viewpoint, men and women disagreed on what constitutes harassment in the case of less severe conduct and in situations involving "peer" harassment.[114] In addition, many people do not consider some social and sexual behaviors harassment, depending on the context in which they occur.[115] These "ambiguous" behaviors include hugs, an arm around someone's shoulder, compliments, joking requests for dates, and sexual jokes and comments.[116] Men rate more ambiguous behaviors to be appro-

understandably worry whether a harasser's conduct is merely a prelude to violent sexual assault. Men, who are rarely victims of sexual assault, may view sexual conduct in a vacuum without a full appreciation of the social setting or the underlying threat of violence that a woman may perceive.

Id.

112. Richard L. Wiener, *Social Analytic Jurisprudence in Sexual Harassment Litigation: The Role of Social Framework and Social Fact*, J. Soc. Issues, Spring 1995, at 167, 174. Women are more likely than men to attribute the cause of sexual harassment to characteristics of the harasser, and less to the character of the harassee. *See* Wiener et al., *supra* note 98, at 264.

113. Charney et al., *supra* note 3, at 11-12. This differential could be explained by men's blindness to and women's perception of the inherent power differential between men and women, even in "peer" situations, due to men's superior physical strength and the history of oppression of women. Older statistics (reported prior to 1982) showed that 80% of harassers were supervisors and only 20% were co-workers. Williams, *supra* note 4, at 74 n.8. More recent statistics show that only 40% of harassers are supervisors, and 60% are peers. Charney et al., *supra* note 3, at 12 nn.22-23 (citing the statistics in the 1981 and 1988 studies of the Merit Systems Protection Board). *See also Sexual Harassment Claims Filed*, J. Acct., Jan. 1997, at 15 (reporting on a study done on 456 companies by the American Management Association which indicated that 49.7% of harassment is done by peers or co-workers, 26.4% by direct supervisors, 17.1% by other supervisors, and 6.8% by customers or vendors). Another recent study found that harassment by supervisors and superiors produced more serious emotional and psychological consequences than did peer harassment. *See* Rebecca A. Thacker & Stephan F. Gohmann, *Emotional and Psychological Consequences of Sexual Harassment: A Descriptive Study*, 130 J. Psychol. 429 (1996); *see also* Roger C. Katz et al., *Effects of Gender and Situation on the Perception of Sexual Harassment*, 34 Sex Roles 35 (1996).

114. Charney et al., *supra* note 3.

115. Wiener, *supra* note 112, at 175.

116. *Id.* This article notes specific factors that influence perceptions of socio-sexual behavior, including the relationship between the parties, the "observer's global evaluation of the incident," the appropriateness of the behavior, the probability that the behavior will be reciprocated, the likelihood of the incident to recur, characteristics of

priate and feel greater comfort with them than do women.[117] Immediacy behaviors—those that function to reduce physical and psychological distance between people—are similar to harassing behaviors in their ambiguity.[118] Immediacy behaviors may create a hostile environment, but they also have been shown to improve the environment; for example, in an academic setting, instructors' immediacy behaviors have been shown to be "positively correlated with students learning."[119] However, researchers have observed that "to the extent that certain harassers may well exploit ambiguity and innuendo to avoid confrontation and rejection, . . . [the] employer's pressures toward clarity of agenda rather than rewards for avoiding it will diminish the difficulties that women face."[120]

Men and women also differ in their interpretations of women's failure to report incidents of sexual harassment. Courts and juries often misinterpret coping behavior as consent.[121] Women indicate that they do not report harassment because they are afraid of hurting their careers, being humiliated, or losing their jobs; they also often believe that nothing can be done to stop the harassment. Unfortunately, "such beliefs are often well founded."[122] Conversely, men are more likely to say that women should handle sexual harassment on their own and not turn to an

---

the observer, and work environment context. *Id.* In addition, there are "significant and widespread" differences in the way in which men and women evaluate socio-sexual behavior, which are influenced by factors such as feeling responsibility for social-sexual conduct at work, outcomes of past interactions, frequency of prior harassment experiences, sex role attitudes, religiosity, internal locus of control, rape myth acceptance, extent to which heterosexual relationships are perceived as adversarial, experience as a victimizer, age, erotophobia versus erotophilia, repression versus sensitization, and social desirability sensitivity. *Id.* at 174-75.

117. Garlick, *supra* note 109, at 152-56.

118. *Id.* at 137 ("Examples of immediacy behaviors in an instructor-student context include maintaining close but appropriate physical distance, touching in an appropriate manner, remaining relaxed, gesturing, spending time with interactants and being vocally expressive.").

119. *Id.*

120. Sarah E. Burns, *Issues in Workplace Sexual Harassment Law and Related Social Science Research*, J. Soc. Issues, Spring 1995, at 193, 197.

121. "[J]uries and appellate courts have consistently construed women's behavior to mean that the harassment was welcome, did not occur, or could not have been that bad—interpretations referred to, respectively as the Slut, Nut, (or So What?) defenses." Fitzgerald et al., *supra* note 4, at 129 (citing Susan Estrich, *Sex at Work*, 43 Stan. L. Rev. 813 (1991)).

122. Fitzgerald et al., *supra* note 4, at 122. *See also* Charney et al., *supra* note 3, at 13.

agent of the employer, and to say that women exaggerate the amount of reported sexual harassment.[123]

B. The Role of Power

Although men and women generally agree that a relationship between a superior and a subordinate tends to increase the likelihood that behavior constitutes harassment, individual supervisors accused of harassment usually deny any intent to harm.[124] Recent research on the effects of power on men may explain the mechanism that produces this "cluelessness," and provide hope that education may help reduce the incidence of sexual harassment.[125]

Persons in positions of power may experience their alleged harassment behavior differently from the way in which their subordinates experience it. A boss who says to his secretary "I find you attractive" might think he is treating her as an equal. The benign interpretation of such behavior is that the boss is blind to his power because he is unaware of his privilege. The critical interpretation is that persons of power who engage in such behavior are abusing their power.[126] The conflicting memories and attributions of intent of the harasser and harassee are "predictable within this framework. They have different memories because they occupy different positions."[127]

In addition, recent research has shown that for some men, power automatically triggers thoughts about sex.[128] Experimental studies have shown that for men likely to sexually harass, the

---

123. Wiener et al., *supra* note 98, at 264.

124. John A. Bargh & Paula Raymond, *The Naive Misuse of Power: Nonconscious Sources of Sexual Harassment*, J. Soc. Issues, Spring 1995, at 87.

125. Conversely, for men who are most likely to harass, sexual desire also may be linked to hostility and/or a desire to dominate women. *See* Fisk & Glick, *supra* note 104, at 99.

126. Kidder et al., *supra* note 76, at 63.

> Just as many Americans and Europeans in the white majority frequently fail to see how they live by "white skin privilege," some professors, physicians, and other institutional actors are blind to their positional privilege. If persons in power fail to see the layers of institutional privilege they embody even as they shift agendas from the professional to the sexual and try to shed their clothing, they fail to understand the threat they represent. They have not shed the power that blinds them.

Id.

127. *Id.* at 63-64.

128. Bargh & Raymond, *supra* note 124, at 85.

WILLAMETTE LAW REVIEW   [Vol. 33:605

idea of power has become habitually associated with the idea of sex.[129] Further, such men rated women who participated in the experiment as more attractive when the concept of power was surreptitiously introduced to these men, by the inclusion of "power" words in the experimental questions they were asked to answer.[130]

The researchers conclude that because the man is unconscious of the link between power and sex, his only explanation for his attraction to the female subordinate is based on factors of which he is aware, such as her physical features and behavior (friendliness, deference, intelligence) or both.[131] On the other hand, the subordinate is aware of the power aspects of the situation, which she sees as an important motivation for her friendly, deferential behavior.[132] Both the man and the woman

> attribute the man's behavior to the features of the situation that are "salient" to them, but these are different. To the man, the salient features . . . are the woman's physical appearance and friendly behavior, but the woman sees instead a man who has power over her [job and future career]. Consequently, she may perceive his sexual advances as an abuse of power, . . . while he perceives his behavior as motivated by personal attraction on his part that seems to be reciprocated [by] her.[133]

The above research may explain why harassers often acknowledge their behavior but do not attach the same meaning or importance to it as the victim does. Reportedly, 75% of harassers "simply don't understand that they are harassers."[134] They do not perceive that they intended to cause distress to the victim; rather, they attribute their actions to another more acceptable motive—e.g., "paying a compliment."[135] Researchers argue that men's lack of awareness

> of the role that their power over a woman plays in their interpretation of her behavior and their own behavior toward her is an important obstacle in changing such behavior. These men will resist the idea that their behavior is an abuse of

---

129. *Id.* at 87.
130. *Id.* at 89.
131. *Id.* at 90.
132. *Id.*
133. *Id.*
134. *Id.* at 87.
135. *Id.*

power because they do not consciously experience any influ-
ence of that power.[136]
This is similar to the "automatic and nonconscious effects of gen-
der and racial stereotypes on judgments and decisions, which
have been documented; here, too, the person often does not in-
tend or consciously experience the operation of the stereotype,
and thus will deny any bias."[137] The researchers hope that, in
the same way that white people have become more aware of the
likelihood that they may have "hidden biases against minority
groups," men can be made aware that

> their feelings of attraction toward their female subordi-
> nates—and their perceptions of the subordinates' attraction
> for them—[may] be subconsciously enhanced by the power
> they hold over those women. Without awareness of this pos-
> sibility, all the good intentions in the world will not help [be-
> cause] . . . the majority of sexual harassers already believe
> that their behavior was motivated by good intentions.[138]

C.   *The Effects of Stereotyping and "Sex Role Spillover"*

Studies of persons in power show that they do not perceive
their subordinates accurately; rather, they rely on stereotypes.[139]
Gender stereotyping is more likely to occur: (a) when there are
few women in a male-dominated environment; (b) when women
move into jobs or professions that were formerly exclusively
male domains; (c) when the workplace is a "sexualized environ-
ment"—i.e., graffiti, pornographic posters and sexual remarks
are tolerated; and (d) when "individuating information" (i.e., in-
formation that helps describe women as unique individuals)
about particular women, as well as evaluative criteria, are ambig-
uous.[140] It takes a great deal of individuating information to dis-
solve stereotypical thinking. This is especially true in the case of
women in occupations that were traditionally dominated by
men.[141]

---

136. *Id.* at 93.

137. *Id.*

138. *Id.* at 93-94.

139. Their subordinates, by contrast, develop "detailed and potentially accurate
impressions of the 'higher-ups' on whom they depend." Kidder et al., *supra* note 76, at
63. *See also* Fisk & Glick, *supra* note 104, at 109.

140. Eugene Borgida et al., *On the Courtroom Use and Misuse of Gender Stere-
otyping Research*, J. Soc. Issues, Spring 1995, at 181, 182.

141. *Id.* at 185, 187-88.

In addition, stereotypes may be used as a method of simplifying complex judgment tasks, regardless of the presence of individuating information. Stereotypes cause or enable people to extrapolate beyond the given data because they assume that all members of the stereotyped group have certain characteristics.[142] For example, women may be stereotyped as dependent or sexual, whereas men may be stereotyped as autonomous and competitive. Thus, the stereotypes influence social judgments because the personal data available will be interpreted differently, depending on the stereotype evoked. Clearly unambiguous information about a woman who rose meteorically to a prestigious management position nonetheless may be construed differently because she is a woman (e.g., "She must have slept her way to the top.") than if the same information were applied to a man (e.g., "He must be highly qualified and ambitious.").[143]

Stereotyping also may occur as a consequence of "sex role spillover," which refers to the tendency for men and women to bring to the workplace the gender roles that guide their interaction elsewhere. According to researchers, because of sex role spillover, "women workers are viewed as sexual with the capacity to elicit sexual conduct from men, and male workers are viewed as 'organizational beings—active, work oriented.'"[144] Stereotypical thinking and sex role spillover tend to cause some men to view women as sex objects rather than as colleagues.[145]

Socio-psychological research can enlighten men and women involved in sexual harassment cases about the genesis, mechanics, and consequences of the harassing behavior. In appropriate cases, women who have been harassed may be able to see their harassers as genuinely lacking in malicious intent; men may be able to understand the effects of their behavior on women, permitting them to apologize without having to see themselves as evil.[146]

While this does not obviate the need for disciplinary action or monetary compensation, it does permit the parties to move toward conciliation and compromise through mutual under-

---

142. *Id.* at 185.
143. *Id.*
144. Wiener, *supra* note 112, at 173.
145. *Id.* at 174.
146. Conversely, in cases involving genuinely hostile men, it may help to remember that fear and insecurity often underlie abusive behavior.

standing. Mediation is the best forum for encouraging the parties to tell their stories, educate themselves and each other about what happened, and reach a solution that will be more satisfying for all concerned.[147]

IV. THE ADVANTAGES OF MEDIATION

Mediation is a voluntary, nonbinding process in which disputing parties try to reach an agreement with the active assistance of a neutral third party—the mediator. Like other species of alternate dispute resolution, mediation allows the parties to tailor a process that will maximize their chances of reaching a satisfactory settlement.[148] Mediation offers the opportunity for each side to see the other's perspective without having to agree with it, and to reach agreement without having to share the same point of view.[149]

In a variety of cases, mediation can be a better means of dispute resolution than traditional litigation, or even arbitration. In sexual harassment cases, the advantages of mediation are even more compelling. Parties in these cases tend to be very emotional, and can cope more easily with mediation than with depositions and trial, which generally are very stressful. Litigators accustomed to operating in the adversarial system may not realize how threatening and unpleasant this environment is for most clients. Sexual harassment cases can be especially difficult because of the need to testify about intimate behavior. In addition, the parties' stories often conflict, and both sides risk adverse credibility judgments, which can be damaging. The plaintiff may be better able to tell her story in the presence of a person who is in a neutral, nonjudgmental role. In sexual harassment cases, the facts often are hotly contested; mediation allows the parties to end the dispute without the need for a credibility determination, which may be destructive to one or both parties.[150]

---

147. For those who desire a more comprehensive understanding of the psychology and dynamics of sexual harassment, as well as additional information on how to prevent it, an excellent source is PETER RUTTER, UNDERSTANDING AND PREVENTING SEXUAL HARASSMENT (Bantam paperback ed. 1997) (originally published in hardcover as SEX, POWER AND BOUNDARIES (1996)).

148. Costello, *supra* note 14, at 20.

149. Stamato, *Sexual Harassment, supra* note 3, at 169.

150. *See* Adler, *supra* note 14.

While mediation can occur at any stage, early mediation is especially desirable in many cases because it allows for a speedier resolution. The necessity of obtaining a "right to sue" letter before proceeding with civil litigation under Title VII or similar state statutes adds additional delay to the usually glacial pace of litigation.[151] All parties stand to save a great deal of stress, time, and expense by settling in mediation.[152] How much they save depends on how early the case is mediated. Mediation better serves the mutual needs of the plaintiff, alleged harasser, and employer for confidentiality and privacy; it allows for creative solutions that are more satisfactory to all the parties.[153] In addition, the parties may choose their mediator based on relevant considerations such as gender, style, age, and knowledge of the substantive law.[154]

Research shows that men and women have different styles of resolving disputes, and that the formality and judgmental aspects of the litigation process better fit a male orientation.[155] Usually female victims simply want the behavior to stop, with no negative consequences to their careers.[156] Women generally are more comfortable with informal processes targeted at ending the problem, rather than finding fault and administering punishment.[157] However, if women suffer secondary injury due to the adversarial nature of litigation, they may respond vindictively; and if the only way of being made whole is through monetary compensation, the price may be quite high.[158] Because it creates a safe environment for telling the story, mediation better meets the needs of women who have been harassed. It allows for catharsis and creative solutions that may be more meaningful for

---

151. *See supra* note 102.

152. Costello, *supra* note 14, at 19-20.

153. *See* Stamato, *Sexual Harassment, supra* note 3, at 169.

154. Adler, *supra* note 14.

155. Charney et al., *supra* note 3, at 13 (citing Stephanie Riger, *Gender Dilemmas in Sexual Harassment Policies and Procedures,* 46 AM. PSYCHOL. 497, 501 (1991); CAROL GILLIGAN, IN A DIFFERENT VOICE (1982)). For more general discussion of differences in the way men and women communicate, see, for example, DEBORAH TANNEN, YOU JUST DON'T UNDERSTAND: MEN AND WOMEN IN CONVERSATION (1990).

156. Charney et al, *supra* note 3, at 13. *See also* Williams, *supra* note 4, at 68.

157. Charney et al., *supra* note 3, at 13 (citing Riger, *supra* note 155; and GILLIGAN, *supra* note 155). *See also* Adler, *supra* note 14.

158. Elizabeth Jubin Fujiwara, *Proving Damages in a Sexual Harassment Case,* 30 TRIAL, April 1994, at 34, 38; Schafran, *supra* note 101, at 14, 22-23.

the plaintiff and at least somewhat less costly for the employer and the harasser.

In addition, mediation can empower the woman who has been harassed, giving her the opportunity to be heard, understood, and treated respectfully. In harassment cases, the plaintiff often has been in a relatively powerless position, and mediation can provide a strong antidote to her powerlessness by allowing her an opportunity to speak. Conversely, in litigation, her feelings of powerlessness often are exacerbated because she must try to be "heard" through her responses to her attorney's questions, which may not be well-designed to let her express herself. She then is subjected to hostile cross-examination. Further, she has no control over the outcome of the case, and is not given an opportunity to seek more meaningful relief, such as an apology from the harasser and the employer or both.[159] Mediation also allows the alleged harasser to explain his perspective and seek forgiveness for unconsciously hurtful acts. Even a defense verdict often does not provide as much satisfaction as does the opportunity to apologize.

In mediation, creative solutions can enhance the likelihood of settlement and can provide greater satisfaction to plaintiffs, often at lower overall cost to defendants. These solutions may include: A written, confidential apology from the harasser and the employer or both, with or without a direct verbal apology during the mediation; the employer's promise to institute sexual harassment training, implement a new policy, or improve or better enforce an existing policy; an offer of transfer or promotion in lieu of job loss; job modification; letters of reference and recommendation; payments in the nature of severance pay; educational funding to provide an opportunity for career change or advancement; acknowledgement of wrongdoing by senior management; employee discipline that "sends a message"; and payment of attorney fees and mediation fees, often in addition to money damages.[160] Moreover, parties sometimes can use media-

---

159. Shereen G. Bingham & Lisa L. Scherer, *Factors Associated with Responses to Sexual Harassment and Satisfaction with Outcome*, 29 SEX ROLES 239 (1993). These researchers found that "although making a formal or informal complaint was not associated with greater satisfaction, talking to the harasser without using aggressive communication strategies increased the likelihood of a satisfactory outcome for the employee." *Id.*

160. This list of creative solutions available in mediation came from the experi-

tion to heal relationships and set the stage for continuing to work together in a positive, supportive manner.

Attorneys sometimes are reluctant to mediate due to fear of revealing "smoking gun" evidence. The experience of many mediators indicates that attorneys often overestimate the value of evidence sought to be concealed; and such evidence can be kept confidential in the mediation if the attorney insists. In many cases, however, such information can be used in mediation to help persuade a party to settle.

There are situations in which mediation cannot work—for example, where the plaintiff seeks only a windfall jury verdict award, where a CEO is willing to spend "millions for defense and not a dime for settlement," or where the mediator or counsel believe that an extreme power imbalance exists that cannot be addressed adequately in mediation.[161] Except in these relatively infrequent situations, however, mediation usually is useful and often successful in resolving the case. Even in cases that do not settle in mediation, the information exchanged and catharsis achieved increase the likelihood of a later settlement prior to trial.[162] In most cases, mediation is by far the best method of resolving a sexual harassment case.[163]

V.   STRATEGIC CONSIDERATIONS IN MEDIATION

Mediations can be useful and successful in many different circumstances. Effective mediators have a variety of personality styles, educational and experiential backgrounds, and approaches. There is no single best approach for any particular case. However, attorneys and mediators should consider factors that may influence the course of the mediation, and attempt to tailor the mediation to help the parties and attorneys feel safe, develop trust, and make a whole-hearted effort to resolve their

---

ence of this author and the mediators and attorneys listed in the acknowledgment at the beginning of the Article (note **).

161. These examples are drawn from the experiences of many mediators and attorneys, as related in interviews with those listed in the acknowledgment at the beginning of this Article, and others who are not named (note **). To preserve confidentiality, the examples, factual scenarios, and tactical suggestions provided in this Article are not credited specifically to particular individuals.

162. *Id.*

163. *See, e.g., supra* note 14.

case. The following are some strategic considerations to assist practitioners.

A. Preliminary Considerations

Before mediation begins, attorneys and parties may make choices that can affect the course and outcome of the mediation. The following are some factors to consider prior to the commencement of mediation:

1. Conflicts of Interest

It is important for attorneys to avoid conflicts of interest. In some cases, mediations have failed because of avoidable conflicts that became painfully apparent during the course of the mediation.[164] To avoid this problem, attorneys considering representing multiple clients need to be aware of potential conflicts. An attorney representing several plaintiffs who were harmed in the same workplace should remember that one client's case may be affected adversely by being "bundled" with the others. For example, where the plaintiff's lawyer represents three sexual harassment clients, one claimant may have a weak case on the facts, another may have great facts but be subject to a unique defense, and the third may have a strong case with egregious facts and no apparent problems. Grouping these cases may be a disservice to the plaintiff with the best case because her case may be diminished by the other less favorable claims. In addition, the plaintiffs may not agree on how to "split the pot," or an initial agreement on a pro-rata split may appear unfair in hindsight.

Similarly, a defense attorney representing both the employer and the alleged harasser may have difficulty doing justice to both clients. For example, in some cases, the harasser maintains innocence and initially has the employer's support; later in the mediation, damning evidence may come to light indicating that the offensive conduct, in fact, did occur. At that point, the employer may want to change its strategy and "throw the harasser to the wolves," creating a painful dilemma for an attorney representing both employer and alleged harasser. Even when the employer and the alleged harasser remain aligned, there can be an issue as to the percentage of the monetary com-

---

164. This is not meant to be an exhaustive discussion of conflicts of interest. Practitioners should consult additional materials for more complete information.

pensation each should pay. In some instances, alleged harassers have sued employers after the resolution of the principal case, claiming unfair treatment in the process.[165] This may be particularly true where the alleged harasser is kept isolated and is not informed of the case's progress, nor allowed to participate in decisions.[166]

These concerns do not mean that attorneys never should represent multiple parties. However, the examples above indicate that attorneys should give some initial thought to potential problems, and make their decisions about whom to represent accordingly.

2. When to Mediate

Often, the best time to mediate is early in the case, perhaps even before litigation is filed. At this point, the parties have not been further polarized, angered, and hurt by the adversarial process; they may have an easier time apologizing, forgiving, agreeing on a settlement, and getting on with their lives. In addition, early mediation saves time and money. Sexual harassment cases can be quite expensive, particularly given the potential for recovery of attorney fees. When cases are resolved without huge litigation costs, defense attorneys have happier clients who may be sources of repeat business. Conversely, plaintiffs' attorneys may realize a higher return on the time they spend on a case when it settles early. Moreover, plaintiffs who settle in mediation are more likely to be satisfied with the overall result, and thus are more likely to recommend their lawyers to others who need representation in similar cases. In addition, both sides benefit from early mediation because it decreases their exposure to publicity and embarrassment.

In some cases, parties or their attorneys may not be ready or able to mediate at an early stage of the case. Sometimes the parties need time to grieve or allow anger and hurt to diminish before resolving the dispute, releasing their anger, and getting on with their lives. In addition, factual discrepancies often require at least some discovery prior to mediation.[167] However,

---

165. Conte, *supra* note 6, at 31.
166. *Id.* at 33.
167. Of course, attorneys need not disclose in mediation any information they desire to keep secret; however, complete development of the facts tends to increase the likelihood that the parties will reach an agreement.

attorneys should be aware that mediation often provides an excellent opportunity for informal discovery, and allows each side time to evaluate the credibility of the opposing party and important witnesses. This informal approach can be far less costly than formal discovery if the case settles in mediation. Formal discovery is often overrated and may not be worth the delay it creates. If the case does not settle in mediation, normally depositions and other discovery still can be accomplished prior to trial.

Even at later stages of the case, mediation can provide a better result for both sides than will a trial and potential appeal. Each side will save time and money, reduce stress, eliminate further uncertainty, and have the opportunity to achieve a more meaningful and satisfying result regardless of when the mediation takes place. If attorneys approach each case with the benefits of mediation in mind, it is more likely that mediation may occur at the earliest and most advantageous time.

3.   *Choosing the Mediator*

One of the obvious advantages of mediation is that the parties may choose the mediator. A party should consider many factors in selecting a mediator, including gender, age, style, experience, and training. Attorneys should weigh all these and other relevant factors, and choose a mediator who inspires their mutual confidence and appears capable of building an environment of trust.

Some believe that female mediators are preferable in sexual harassment cases.[168] However, experience demonstrates that male mediators can be equally effective. The mediator's gender may be more significant in some cases than in others. When a female plaintiff has been frightened or seriously aggrieved, and in cases involving real or perceived power imbalances, a female mediator may be a better choice; the plaintiff may more easily feel understood by and have confidence in a woman. Conversely, if the female plaintiff feels the need for the appearance of greater physical security in the mediation itself, a male mediator may appear better able to control the process and enforce civility. In cases where the plaintiff is a fairly strong woman and

---

168. Barry Winograd, *Men as Mediators in Cases of Sexual Harassment*, 50 DISP. RESOL. J., June 1995, at 40.

the alleged harasser is nervous and defensive, a male mediator may be more effective.

The degree of severity of the harassment, and the relationship between harasser and harassee, also may affect mediator gender preference. Studies have shown that women are more likely to complain about sexual harassment, feel more confident, and are less prone to unwarranted self-blame in situations where the conduct was more egregious and when the harasser is her boss.[169] Moreover, in such cases, the harasser may be more defensive. In addition, if we assume that the ultimate root of abusive behavior is fear and insecurity, we might conclude that the most outrageous harassers are also the most fearful and insecure. In such cases, perhaps involving quid pro quo harassment and extreme behavior or both, a male mediator may be more able to help support the harasser and allow him to admit his behavior and apologize. Conversely, in cases of more ambiguous behavior, and in hostile environment situations, women may be more likely to blame themselves. In these cases, female mediators may be more sensitive to the plaintiff's perspective, and better able to validate her sense of violation. Cases in which a male is harassed are fairly rare,[170] so sufficient research and experiential basis does not exist to permit informed assumptions about how mediator gender might affect the mediation; however, some of the same considerations described above might apply by analogy in such cases.

The mediator's style also is an important consideration. In some cases, it might be more important to choose a mediator who is more gentle, or relationship-oriented, or whose style is more facilitative or interest-based. In other cases, a directive style or evaluative "settlement conference" approach may be more useful.[171] In reality, many cases can be resolved satisfacto-

---

169. Sarah Barton Samoluk & Grace M.H. Pretty, *The Impact of Sexual Harassment Simulations on Women's Thoughts and Feelings*, 30 SEX ROLES 679, 694 (1994).

170. *See supra* note 4 and accompanying text.

171. Mediators with a "facilitative" style focus on helping the parties communicate more effectively and enhancing their ability to resolve the dispute, without giving opinions about the merits or value of either side's case. They may, however, draw out evaluative information by questioning the attorneys (usually in separate caucus) about their view of the likelihood of certain outcomes. "Evaluative" mediators, conversely, offer their legal opinions and analysis. "Interest-based" mediation refers to the efforts made to identify all the interests and needs of the parties, including those normally not addressed in litigation—such as emotional concerns, values, intangible goals, etc.

rily with mediators who have a range of styles; many mediators have flexible styles and can adapt their approaches to suit the needs of a particular case.

Depending on the ages of the parties and their attorneys, the age of the mediator also may be relevant. When a young woman claims she was harassed by an older male, she may have difficulty trusting an older male mediator. Conversely, an older female plaintiff may not believe that a younger mediator of either sex can understand her experience. In cases where the alleged harasser is very nervous and defensive, it may be more important to choose a mediator in his age group.

A mediator's experience, training, and knowledge of sexual harassment law also may be important. While in many cases the mediator's competence in managing the process is more relevant than his or her knowledge of substantive sexual harassment law, sometimes the parties or attorneys may have greater confidence in a substantively knowledgeable mediator. Attorneys should discuss their needs in these matters with proposed mediators.

Finally, attorneys should remember to consider the big picture, rather than focusing solely on their client's separate interest. The success of the mediation depends not only on one side's comfort zone, but on whether the overall needs and interests of both sides are met. In many cases, an attorney may best serve his or her client's interests by choosing a mediator who can inspire the opposing party's trust, particularly when the other party apparently has a narrower range of comfort. Attorneys can enhance the possibility of settlement by sharing information about their respective clients' needs and attempting to accommodate both parties.

4. Who Should Attend the Mediation?

It can be crucial to the success of the mediation to ensure

Many, if not most, mediators use an "interest-based" style, but some still may come from a "position-based" style that focuses exclusively, or nearly so, on the legal issues in the case and the parties' positions on those issues. For more in-depth information on mediation theory, see, for example, ROBERT A. BARUCH BUSH & JOSEPH P. FOLGER, THE PROMISE OF MEDIATION (1994); ROGER FISHER & WILLIAM URY, GETTING TO YES: NEGOTIATING AGREEMENT WITHOUT GIVING IN (1981); JAY FOLBERG & ANN TAYLOR, MEDIATION: A COMPREHENSIVE GUIDE TO RESOLVING CONFLICTS WITHOUT LITIGATION (1984); CHRISTOPHER MOORE, THE MEDIATION PROCESS: PRACTICAL STRATEGIES FOR RESOLVING CONFLICTS (1986).

that all important "players" attend and participate. These important participants may include some or all of the following: claimant, alleged harasser, supervisor, management representative, insurance claims representative, and significant witnesses; support people such as spouses, significant others, friends, family members, and therapists. Sometimes the absence of an important person makes resolution more difficult. For example, in one case, the plaintiff would not accept a monetary settlement until she received an apology from a management representative and the assurance that the president of the company knew of her claim and had authorized the settlement offer personally.[172] It is critically important that the plaintiff feel heard, understood, and respected. In cases where it is possible for the parties to talk directly with one another, the victim's awareness that the alleged harasser and the employer's representative have heard her may facilitate settlement. It may be equally important for the man accused of harassment to communicate his perspective directly to his co-worker or subordinate, including subjectively benign intentions, and to offer his apology. Even in the "he said, she said" cases in which the facts remain in dispute, mediation can result in a settlement when each side has the opportunity to tell its story even without the parties having to agree on the facts. Having all important players at the mediation greatly facilitates this process.

When spouses and significant others attend the mediation, the attorneys and mediator may need to make tactical decisions about whether, how, and when to include them. In some cases, acknowledgement of improper conduct and complicity may be important to resolution; but the presence of a spouse or significant other who has a personal stake in his or her partner's "innocence" may interfere with or prevent this. In such cases, it may be helpful to arrange for the harasser and harassee to talk alone with the mediator, or to excuse all support people from a separate caucus or joint session, in order to fully develop the facts and permit the parties to make necessary acknowledgements and apologies. In one case, after the initial storytelling, the mediator suspected that at one time there had been a consensual, voluntary relationship between the parties that had soured, creating bad feelings on both sides and resulting in an accusation of har-

---

172. *See supra* note 161.

assment.[173] In that situation, the mediator arranged for the parties to talk alone in the presence of the mediator, where they were able to acknowledge their mutual involvement and apologize for the difficulties they had caused each other.[174] After that meeting, the mediator sent each party off to talk with his or her attorney; eventually, after additional negotiation, the parties reached a settlement.[175] The case probably would not have settled at the mediation without the mutual understanding and exchange of apologies. The mediator took a risk in bringing the parties together; but the catharsis, closure, and resulting willingness to consider a reasonable settlement probably would not have happened without doing so.

5. Preparing for the Mediation

Prior to the mediation, attorneys can increase the likelihood of a satisfactory resolution by preparing themselves and their clients. Attorneys need to "shift gears" from an adversarial mode to a collaborative one, particularly in cases fraught with acrimony between counsel as well as between the parties. The legal system is set up as a search for the objective "truth." Therefore, lawyers must adopt a mindset in which reasonable people legitimately may have different perspectives. This orientation may permit the opportunity for a resolution by allowing a face-saving view of the dispute as the product of differences in perspective. The attorneys should embody and model civility, objectivity, a conciliatory attitude, and a willingness to listen to the other side with a relatively open mind. This will be easier if the attorneys take the time to understand the potential bases for a legitimate difference in perspective between the parties[176] and, in appropriate cases, educate their clients about this information. Defense attorneys often need to remember to be gentle in mediation to avoid retraumatizing the plaintiff. This may require a conscious mental shift for lawyers who have an aggressive, adversarial approach to litigation.[177] Even when the defense attorney has a

---

173. *See supra* note 161.
174. *See supra* note 161.
175. *See supra* note 161.
176. *See supra* Part III.
177. For more discussion on how lawyers can best serve their clients by "turning down the heat," see Charles Guittard, *Detoxifying Sensitive Cases*, LEGAL TIMES, July 15, 1996, at 13.

fairly jaded view of the plaintiff, courtesy, kindness, and gentleness can facilitate settlement.

Attorneys can prepare their clients to generate creative solutions in mediation by asking questions that will help them identify their needs and what they might offer to the other side, such as: "What do you need to have happen in order to feel better?" "What works?" "What would it take for you to be able to resolve this and get on with your life?" "What would be fair?" "What is necessary to correct the imbalance of power?" "Would an apology help?" "Is it important to heal this relationship?" "Even if you cannot agree with the other side's version of the facts, is there something you can ask or offer that might help both sides to resolve this dispute?" These and similar questions can be invaluable in preparing the parties for mediation; they help the parties explore possible creative solutions and identify intangible needs that must be satisfied before the case can be resolved. The attorneys also can help prepare the mediator by conveying this information to her or him. Focusing on pragmatic solutions, emotional needs, and intangible interests is more likely to produce a satisfactory outcome than honing the client's version of the facts and reinforcing the client's position.

B. The Mediation Process

The format and process of mediation varies, depending on the mediator's style and the parties' needs. Many factors influence the flow of the mediation. Some involve strategic choices, and some are essential elements in any mediation.

1. Joint Session versus Shuttle Diplomacy

Mediators have different views of how a sexual harassment mediation should be structured. Should it begin with a joint session, in which the mediator introduces and explains the process and then encourages each side to tell its story in the presence of the other? Or are sexual harassment cases so fraught with emotional tension and difficulty that shuttle diplomacy is the only way to proceed?

Despite the risk, many mediators feel that starting with a joint session is extremely important to the outcome of the mediation. They view the joint session as an invaluable way to set the tone for mutual respect and understanding. Providing a safe place for the parties to share their stories with one another can

be crucial to each side's ability to see the other party's legitimately different perspective. These mediators feel that, even in cases where the facts will remain in dispute, the likelihood of reaching a resolution is much greater when each side is heard, understood, and respected by the other.

To help the participants see each other as individuals rather than combatants, the mediator may start the mediation with fairly extensive introductions of both the parties and their attorneys. This does not take long, but puts a human face on the process. At the beginning of the joint session, the mediator explains the process and answers any questions. By establishing the ground rules—such as requiring that only one person speak at a time and that everyone listen with respect even when they disagree—the mediator creates a safe environment for all participants. Starting with a joint session allows everyone to begin "on the same page," enhancing trust. The mediator's calm, matter-of-fact way of dealing with potentially explosive subject matter sets an example that encourages the parties to relate their stories with feeling but without histrionics. The assumption that all participants are capable of civil discourse tends to be a self-fulfilling concept, in the same way that other human behavior—whether of children or adults—tends to rise or sink to expectations. The joint session also creates an opportunity to provide educational information to all parties—such as the fact that research has demonstrated that men and women simply have different perspectives and may disagree about what constitutes sexual harassment. This allows the parties to step back from their own dispute, and realize that this dilemma is not unique to them and is a cultural problem, rather than strictly the product of the other side's malice or hypersensitivity.

When mediators start with a joint session, they usually try to create a collaborative, educational environment rather than continue the adversarial interactions that have preceded the mediation.[178] For example, rather than having the attorneys start with opening statements, the mediator may ask the plaintiff to tell her story first, and then have her attorney briefly outline the issues objectively. Then the alleged harasser and his attorney follow the same procedure. The focus is not on persuading, but rather on ensuring that everyone understand's each side's story. Expe-

---

178. *See supra* note 161.

rience indicates that even parties who are nervous can speak with conviction, and can be articulate, if not eloquent. Just being able to tell one's story is empowering and helps correct any actual or apparent power imbalance. Also, having parties talk about their suffering allows for the development of empathy. Attorneys who are skeptical of the benefit of having the parties speak for themselves may appreciate the opportunity to learn how the client and the opposing party present themselves under pressure. There is no need in mediation to reach an agreement on "the true facts," nor is it necessary to decide which party is more credible. Sometimes the parties can agree on what happened, and sometimes they cannot; the important point is that each side is heard, understood, and respected. Experience indicates that when the parties participate actively, the mediation tends go more quickly and smoothly.[179]

Some mediators use a "shuttle diplomacy" model in all sexual harassment cases because they view these cases as too volatile to risk a joint session.[180] These mediators feel that, in a joint session, the plaintiff may be retraumatized or that one or both parties will be unable to listen without becoming explosively angry, which might doom the mediation.[181] In shuttle diplomacy cases where the mediator prefers to prevent direct contact betweeen the parties, the mediator may have to take extreme steps to ensure separation, such as using conference rooms on two different floors of a building and having the parties arrive and leave at different times in order to avoid meeting in the elevator.

However, even mediators who prefer this approach sometimes get the parties together. This may be done to allow a direct apology in the hope of breaking an impasse, or to clarify complex facts. In some cases, the plaintiff may resist meeting with the alleged harasser but is willing to speak directly with a management representative or insurance claims person. This approach offers her the opportunity to experience the catharsis and closure that telling her story and feeling heard and respected affords, without feeling threatened or frightened. In this scenario,

---

179. *See supra* note 161.
180. *See supra* note 161.
181. *See supra* note 161.

the mediation process offers some benefits of a joint session, at a reduced risk.

2. *Use of Separate Caucus*

Most frequently, even in mediations that begin with a joint session, the mediator subsequently will separate the parties and utilize separate caucuses. In this nonthreatening and confidential environment, the mediator can encourage complete development of the facts, permit emotional venting and catharsis, provide reality-testing, and allow discussion of what each side could offer the other.[182]

Although the parties may develop the facts fairly completely during the joint session, additional details, and often crucial nuances, usually come to light only when the mediator spends time alone with each side. The mediator may help the parties understand why sharing this information with the other side is useful or why keeping it confidential is preferable. At a minimum, it usually is important for the mediator to have as complete an understanding of all the facts and concerns of each side as possible. Separate caucuses provide an opportunity for the mediator to draw out sensitive information not revealed in the joint session.[183]

To ensure a safe environment in the joint session, the mediator must require that the parties remain civil and discuss their emotions in a nonabusive manner. It may be important to provide an opportunity for each side to express strong emotions in the safety of the separate caucus. Each side needs to feel heard and understood and have their genuine emotions validated, regardless of how they have treated each other prior to the mediation. This cathartic experience often must occur before the parties can go of the dispute and work on a pragmatic resolution.

At some point in the separate caucuses, mediators become the "agent of reality." The mediator may ask questions such as: "How do you think a jury might view that?" "Do you think the judge is going to allow that evidence to come in?" "Do you think there is a possibility that jurors who are forced to choose whose story is true might favor the other side?" The focus is not on finding out who is lying, but, rather, on examining what an

---

182. *See supra* note 161.
183. *See supra* note 161.

objective fact-finder might do with the stories presented. Often, parties have legitimate strong feelings but nonetheless can concede weaknesses in their case. This can begin to help them move into an objective mode and consider reasons to be pragmatic because they realize that there is a chance that they could lose the case at trial. In addition, in separate caucus, the mediator can discuss with the attorneys in the presence of their clients the risks, costs, and stress involved in continuing litigation.

After the mediator and participants in separate caucus have fully developed the facts, provided an opportunity for and experienced emotional catharsis, and engaged in reality-testing, the parties will begin the task of generating possible elements of a resolution. At this point, the mediator will ask the attorneys and parties to discuss creative solutions, and the mediator can make suggestions if they have difficulty with this task.

Throughout the separate caucus stage of the mediation, the mediator moves between the parties, conveying information, facilitating expression of emotion, keeping the parties working toward the goal of settlement, presenting tentative offers, and performing any other useful functions. At some point during the caucus stage, or after an agreement has been reached, the mediator may reunite the parties to convey an apology and expression of forgiveness, exchange additional information, clarify complex facts, or confirm verbally the terms of their agreement.

3. *Additional Considerations*

Other details are important in the mediation. Even minor power imbalances need to be addressed and dealt with in order for the mediation to succeed. Where an ongoing relationship is involved, additional mediation sessions may be necessary to keep things on track. Due to their complexity and emotional volatility, sexual harassment mediations can be more time-consuming than other kinds of mediations; parties and attorneys need to be prepared to commit the time necessary to optimize the likelihood of a successful outcome. Often parties who are nervous do not eat before the mediation; keeping everyone adequately fueled can be a real concern.

Finally, the parties and attorneys need to remain flexible throughout the mediation—and release an attachment to a particular outcome—in order to evaluate each proposal from a fresh perspective and maintain the ability to generate new ideas.

This may sound a bit Zen-like, but actually it is quite pragmatic. It has been said that the person with the most flexibility has the most power in mediation.[184] This is because people who become rigid and fixated on particular points have fewer options for a successful resolution than do those who remain open-minded and capable of considering all offers and generating creative ideas.

VI. CONCLUSION

Although we have made significant progress as a society, sexual harassment is a problem that probably will persist for some time. The change required in order to eliminate it is fundamental; resistance to such change is strong, even in those who have received education and want to change. Mediators and attorneys can help resolve individual cases, as well as expedite fundamental social change, by educating parties about the perceptual differences between men and women that this Article describes.[185] Experience and research demonstrate that mediation is a much better process for resolving sexual harassment claims than litigation or arbitration; mediation works best when attorneys and participants are well-prepared and committed to the effort required for successful resolution of their disputes.

---

184. The author first heard this expressed by James Melamed and Robert Benjamin during a mediation course they presented in Portland, Oregon in May-June 1993.

185. Preventive training is beyond the scope of this Article, but a number of other articles and books address this topic. *See, e.g.,* RUTTER, *supra* note 147; Joe Schumacher & Judy Fester, *Practical Ways to Address Workplace SXH Problems,* PUB. MGMT., July 1996, at 19; Dorraine A. Larison & Mary E. Olk, *Sexual Harassment Awareness Training: It's Not the Boogie Monster,* 72 N.D. L. REV. 387 (1996).